INTERNATIONAL ECONOMIC REVIEW

September, 1960
Vol. 1, No. 3

CONTENTS

Some Examples of Global Instability of the Competitive Equilibrium *H. Scarf* 157

Simultaneous Least Squares : A Distribution Free Method of Equation System Structure Estimation *T. M. Brown* 173

A Note on the Consistency of Simultaneous Least Squares Estimation *M. Nakamura* 192

Problems in Production-Investment Planning over Time *V. L. Smith* 198

Sir Roy F. Harrod on Internal and External Balance *M. Tatemoto* 217

Tariffs and Trade in a Two-Good World *W. M. Gorman* 223

The Long-Run Development of National Income in Germany : A Review Article *B. Dietrichs and R. Jochimsen* 230

Published by
THE KANSAI ECONOMIC FEDERATION, OSAKA

INTERNATIONAL ECONOMIC REVIEW

Published by the Kansai Economic Federation with
the Collaboration of the Wharton School of Finance and Commerce,
University of Pennsylvania, and the Institute of Social and
Economic Research, Osaka University.

CHIEF EDITOR

L. R. KLEIN
Department of Economics,
Wharton School of Finance
and Commerce, University
of Pennsylvania,
Philadelphia 4, Pa.,
U.S.A.

CO-EDITOR

M. MORISHIMA
The Institute of Social
and Economic Research,
Osaka University,
Toyonaka, Osaka,
Japan

ASSOCIATE EDITORS:

M. ALLAIS (Ecole Nationale Supérieure des Mines de Paris, France)
T. M. BROWN (Royal Military College, Canada)
G. DEBREU (Yale University, U.S.A.)
R. M. GOODWIN (Cambridge University, U.K.)
F. H. HAHN (University of Birmingham, U.K.)
B. HANSEN (Konjunkturinstitutet, Sweden)
L. HURWICZ (University of Minnesota, U.S.A.)
D. PATINKIN (The Hebrew University, Israel)
P. N. RASMUSSEN (University of Copenhagen, Denmark)
N. S. R. SASTRY (Reserve Bank of India, India)
R. H. STROTZ (Northwestern University, U.S.A.)
T. W. SWAN (Australian National University, Australia)
H. THEIL (Netherlands School of Economics, Holland)

BUSINESS OFFICE OF THE INTERNATIONAL ECONOMIC REVIEW: Kansai Economic Federation,
Osaka Building, Soze-cho, Kita-ku, Osaka, Japan.

Copyright © 1960 by Kansai Economic Federation.

Printed by Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.),
No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

INTERNATIONAL ECONOMIC REVIEW

September, 1960
Vol. 1, No. 3

SOME EXAMPLES OF GLOBAL INSTABILITY OF THE COMPETITIVE EQUILIBRIUM*

BY HERBERT SCARF[1]

IN THIS PAPER we shall consider the problem of stability of the competitive equilibrium. The market demand functions are sums of individual demand functions obtained directly by utility maximization. The rate of change of the price of each commodity is assumed to be proportional to the excess market demand for that commodity. A number of examples are given for which the motion of the prices is globally unstable in the sense that starting from any set of prices other than equilibrium, the prices oscillate without tending towards equilibrium.

1. INTRODUCTION

The problem of stability of the competitive equilibrium is described in [3], and we shall content ourselves with a review[2]. Several individuals with utility functions $U_i(x_1, \cdots, x_n)$ for the same commodities are engaged in the trade of these commodities. Each individual begins the trading with an initial endowment of goods, say $I_1^i, I_2^i, \cdots, I_n^i$, for the ith individual. An initial vector of prices $p_1, \cdots, p_n$ is announced, and each consumer then determines his demand for all of the commodities by the usual procedure of maximizing his utility function, subject to the constraint that his expenditure shall not exceed the value of his initial endowment of goods at the stated price vector.

For each commodity the sum of the individuals' demand functions

* Manuscript received January 21, 1960.

[1] Research undertaken by Stanford University under Contract Nonr-225(28), NR 047-019 with the Office of Naval Research and completed at the Cowles Foundation for Research in Economics under Contract Nonr-358(01), NR 047-066 with the Office of Naval Research. I would like to thank K. J. Arrow, L. Hurwicz, and H. Uzawa for a number of stimulating conversations.

[2] The reader may wish to consult the excellent bibliography given in [3] and also in [1].

minus the sum of the initial endowments of that commodity is called the market excess demand for the commodity; in this paper it will be denoted by $f_i(p_1, \cdots, p_n)$, where the subscript refers to the commodity in question. (These functions are homogeneous of degree zero and satisfy the Walras Law $\sum p_i f_i \equiv 0$.) Excess demand functions for individuals will always be denoted by x_i.

An equilibrium price vector is, of course, a vector of prices for which all of the excess demand functions vanish, and recent work in this area [2, 8] has shown that under suitable regularity conditions equilibrium prices will always exist. (Because of the homogeneity of the demand functions, any positive multiple of an equilibrium is again an equilibrium, and in this sense we should speak of an equilibrium ray. There may be several such rays.) The stability problem, on the other hand, is less concerned with the existence of equilibrium and more with the question of what happens to the prices if initially they are different from the equilibrium.

There is nothing in the model described so far which enables us to compute the motion of prices if we are not at an equilibrium point. For this we need some specific assumptions on the price adjustment process, that is, the procedure by which prices may be expected to change if we are away from an equilibrium price. The intuitive notion that an excess of supply over demand should result in a decrease in price, and an excess of demand should result in an increase in price, has been formalized mathematically in [7] and [3] by the statement that

$$\frac{dp_i}{dt} = H_i[f_i(p_1, \cdots, p_n)] ,$$

where H_i is a sufficiently regular sign-preserving function of its argument. (These equations are to hold if all $p_i \geq 0$.) In order to be concrete, in this paper we shall generally take the functions H to be equal to their respective arguments, so that for each commodity the rate of change of price is equal to the excess demand.

Now let us turn our attention to the problem of stability with this type of adjustment process. Early work in this area [5, 6] tends to emphasize what might be called "local" stability; the initial prices are assumed to be close to some equilibrium point, and an analysis is made of whether there is a tendency to converge to the equilibrium point, depart from the equilibrium point, or perhaps even a tendency to more complex types of behavior. The "local" analysis proceeds by means of the linear terms of the Taylor series expansion about the equilibrium

point, thus converting the problem to a linear differential system with constant coefficients. It is possible in this type of analysis to obtain many examples of completely unstable equilibria (though it should be mentioned that there is an uncomfortable tendency for examples to be produced without any consideration of their origins as market demand functions derived by the summation of individual demand functions). The local analysis, however, is somewhat unsatisfactory, in that it is quite possible for other equilibrium points to exist, and the system cannot be said to be unstable without examining whether the prices tend to another equilibrium point.

This consideration leads naturally to the problem of stability in the "global" sense [3], which is concerned with the solution of the differential equations based on the price adjustment mechanism for an arbitrary initial set of prices. If the solution of the differential equations approaches *some* equilibrium point as time becomes infinite, then we have global stability. Clearly it is quite possible for there to be several equilibrium points, none of which is completely stable from the local point of view (that is, attracts all neighboring points) and where the system, in its entirety, is globally stable.

The problem of global stability, as it has been described above, has been discussed by several authors. Though alternate methods of proof have been found, the results do not go essentially beyond those obtained in [1]. There are, first of all, several special cases which can be mentioned: global stability is known if there is a single consumer, or if there are only two goods, or if it happens that at some equilibrium point there is no trade. There is also an unpublished result of the author's showing that global stability is known if all of the consumers have the same homogeneous utility function, but with different initial distributions of stock. Aside from these very special cases, the most important result is that global stability will occur if all of the goods are gross substitutes; mathematically, this means $\partial f_i/\partial p_j > 0$ if $i \neq j$. Apart from this, very little else has been found; no assumptions substantially different from gross substitutability have been shown to imply stability, and up to the present no examples of instability have been produced.

This paper presents a series of examples, all derived from utility maximization, which are globally unstable. The examples given here involve three consumers and three goods, but the techniques may be extended to either more consumers or more goods. The examples all involve a rather simple relationship between the utility functions of the three consumers, namely

$$
\begin{aligned}
U_2(x_1, x_2, x_3) &= U_1(x_3, x_1, x_2) , \\
U_3(x_1, x_2, x_3) &= U_2(x_3, x_1, x_2) ,
\end{aligned}
\tag{1}
$$

and a corresponding cyclic permutation of the initial endowments. Specifically, we will assume that if the initial endowments of the first consumer are (I, J, K) for the first, second, and third goods respectively, then those of the second consumer are (K, I, J) and those of the third consumer (J, K, I). This relationship is introduced to simplify the calculations, and, as the method will show, it may be relaxed.

It is an easy consequence of utility maximization that for the price adjustment process we have selected (rate of change of price equal to excess demand) the motion of the prices will always be constrained to the sphere $p_1^2 + p_2^2 + p_3^2 = \text{constant}$. In all of our examples there will be an equilibrium point at $(p_1 = p_2 = p_3)$, which, except for the first example, will be locally completely unstable (there is a small region around this equilibrium point such that if the initial price is in the region, the prices will eventually leave this region and stay out). In the first example the motion of the prices will be in concentric curves about the equilibrium point, whereas the more complex examples will give rise to limit cycles.

As we shall see from the examples of Section 3, instability does not depend on a delicate assignment of values of initial stocks or parameters in the utility functions. Though it is difficult to characterize precisely those markets which are unstable, it seems clear that instability is a relatively common phenomenon.

What are the implications of these examples? It seems to me that there are several possible interpretations that *might* be made.

1. One possible interpretation is that the model is substantially realistic and that instabilities of the type described in this paper could possibly occur. An even more presumptuous interpretation along these lines is that instability is responsible for some aspects of the business cycle, though for this sort of interpretation it would seem advisable to produce examples of instability with a model of a complete economy rather than a pure trade model alone, and this may be more difficult to do.

2. Another possible interpretation is that the price adjustment process postulated above is not correct. This view can, of course, be held without any reference to the question of stability. An argument for this position is that in one sense or another we are considering a dynamic process, and yet nowhere do the simplest dynamic considerations such as saving, interest, etc. appear in the model.

3. As a final interpretation it might be argued that the types and

diversities of complementarities exhibited in this paper do not appear in reality, and that only relatively similar utility functions should be postulated, and also that some restrictions should be placed on the distribution of initial holdings. This view may be substantiated by the known fact that if all of the individuals are identical in both their utility functions and initial holdings, then global stability obtains.

2. A VERY SIMPLE EXAMPLE OF INSTABILITY

In this section we shall describe a very simple example which leads to instability. It will be seen that this example is quite an extreme one. In the next section we shall describe a number of additional examples which, while somewhat more complex, do not have the disagreeable features of the present one.

Let the utility function of the first consumer be $U_1(x_1, x_2, x_3) = \min(x_1, x_2)$ and his initial endowments $(1, 0, 0)$. This consumer has no desire for the third good and his indifference curves for the first two goods are of the form:

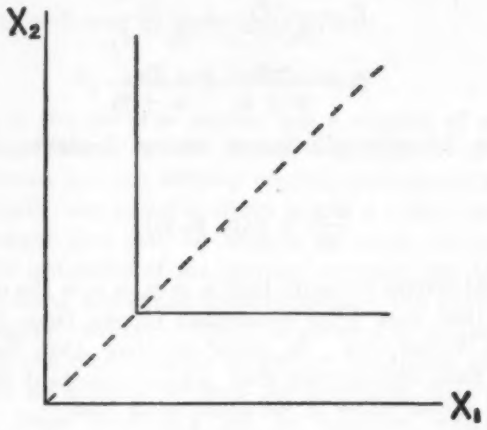

FIGURE 1

For any income M the same quantity will be demanded of goods one and two and therefore the demand functions are

$$y_1(p_1, p_2, p_3, M) = \frac{M}{p_1 + p_2},$$

$$y_2(p_1, p_2, p_3, M) = \frac{M}{p_1 + p_2},$$

$$y_3(p_1, p_2, p_3, M) = 0.$$

Now the income of the first consumer is derived from his endowment of a single unit of good one, so that $M = p_1$ and therefore the excess demand functions of the first consumer (supply is being subtracted off now) are given by

$$x_1 = \frac{-p_2}{p_1 + p_2},$$

$$x_2 = \frac{p_1}{p_1 + p_2},$$

$$x_3 = 0.$$

The excess demand functions of the second consumer are obtained by a cyclic permutation of all subscripts, $1 \to 2, 2 \to 3, 3 \to 1$, and once more for the third consumer. If we add these together, the market excess demand functions are given by

$$f_1 = \frac{-p_2}{p_1 + p_2} + \frac{p_3}{p_3 + p_1},$$

$$f_2 = \frac{-p_3}{p_2 + p_3} + \frac{p_1}{p_1 + p_2},$$

$$f_3 = \frac{-p_1}{p_3 + p_1} + \frac{p_2}{p_2 + p_3},$$

and of course the price adjustment process leads to the differential equations

$$\frac{dp_i}{dt} = f_i(p_1, p_2, p_3). \tag{2}$$

It is a trivial matter to verify that $p_1 = p_2 = p_3$ is the only equilibrium point. The fact that $\sum p_i^2 = \text{constant}$ follows from $\sum p_i dp_i/dt = 0$, which is the Walras Law. In order to show that the solutions are unstable we shall demonstrate that $p_1p_2p_3 = \text{constant}$ for any solution of (2). This would follow if we could show that

$$f_1p_2p_3 + f_2p_1p_3 + f_3p_1p_2 \tag{3}$$

equals zero, but (3) is equal to

$$\frac{p_3(p_1^2 - p_2^2)}{p_1 + p_2} + \frac{p_2(p_3^2 - p_1^2)}{p_2 + p_1} + \frac{p_1(p_2^2 - p_3^2)}{p_3 + p_3}$$

$$= p_3(p_1 - p_2) + p_2(p_3 - p_1) + p_1(p_2 - p_3) = 0.$$

That this implies instability is clear. Let the initial prices be chosen such that $p_1^2 + p_2^2 + p_3^2 = 3$, so that the intersection of the equilibrium ray and this sphere is $(1, 1, 1)$. The value of $p_1p_2p_3$ at equilibrium is

one, and therefore if the initial price gives a value different from *one* to $p_1p_2p_3$ we never reach equilibrium. It should be remarked that the maximum of $p_1p_2p_3$ subject to the constraint $\sum p_i^2 = 3$ is actually *one*, so that if the initial position is anything other than (1, 1, 1), the path is completely unstable.

Occasionally the following price adjustment process is discussed. One of the goods is singled out and its price kept constant; the remaining goods are meant to vary according to the differential equations given above. If, in our case, we put $p_3 \equiv 1$, then we are led to the system

$$\frac{dp_1}{dt} = \frac{-p_2}{p_1 + p_2} + \frac{1}{1 + p_1},$$

$$\frac{dp_2}{dt} = \frac{-1}{p_2 + 1} + \frac{p_1}{p_1 + p_2}.$$

Routine calculations show that for this system

$$p_1p_2e^{-\frac{1}{2}(p_1^2+p_2^2)} = \text{const.},$$

and this is again sufficient to show instability.

3. A CLASS OF EXAMPLES

The example of the previous section has a number of special properties. All of the Slutsky terms $[\partial y_i/\partial p_j + y_j(\partial y_i/\partial M)]$ are zero, the indifference surfaces are not *strictly* convex, and certainly not differentiable, and, finally, the initial holdings are of a rather extreme type. It might be thought that one, or several, of these properties is responsible for the instability of the previous example, and that stability would return if these properties were removed. The examples of the present section show, however, that it is quite easy to obtain instability with none of the objectionable properties mentioned above.

We shall make an attempt to keep the reasoning relatively general in this section. Some conditions will be described which imply instability, and we shall demonstrate by specific examples that these conditions may be satisfied.

As was mentioned in the introduction [see equation (1)], the utility functions of the three consumers will be obtained by a cyclic permutation of the goods and the initial endowments. This means that if the *excess* demand functions of the first consumer are represented by

$$x_1(p_1, p_2, p_3), \quad x_2(p_1, p_2, p_3), \quad x_3(p_1, p_2, p_3),$$

then the excess demand functions of the second consumer for the first, second, and third goods respectively will be given by

$$x_3(p_2, p_3, p_1), \quad x_1(p_2, p_3, p_1), \quad x_2(p_2, p_3, p_1),$$

and those of the third consumer by

$$x_3(p_3, p_1, p_2), \quad x_3(p_3, p_1, p_2), \quad x_1(p_3, p_1, p_2).$$

(Continuity and differentiability properties of these functions will be assumed whenever necessary.)

The market excess demand functions are obtained by the summation of individual demands and are therefore given by:

$$\begin{aligned} f_1(p_1, p_2, p_3) &= x_1(p_1, p_2, p_3) + x_3(p_2, p_3, p_1) + x_2(p_3, p_1, p_2), \\ f_2(p_1, p_2, p_3) &= x_2(p_1, p_2, p_3) + x_1(p_2, p_3, p_1) + x_3(p_3, p_1, p_2), \\ f_3(p_1, p_2, p_3) &= x_3(p_1, p_2, p_3) + x_2(p_2, p_3, p_1) + x_1(p_3, p_1, p_2). \end{aligned} \tag{4}$$

The differential equations of stability are, of course, given by $dp_i/dt = f_i(p_1, p_2, p_3)$, when all prices are non-negative.

As we mentioned previously, the Walras Law ($\sum p_i x_i \equiv 0$ for each consumer, and its consequence, $\sum p_i f_i \equiv 0$) implies that $d(\sum p_i^2)/dt = 0$, so that the motion of the prices will always be on that sphere with center at the origin which passes through the initial price vector. Let us assume that the initial prices have been selected such that $\sum p_i^2 = 3$. When we speak of equilibrium prices we shall mean the intersection of the equilibrium ray and this sphere.

LEMMA 1. *The price vector* (1, 1, 1) *is an equilibrium price.*

This is immediately obvious by an application of the Walras Law to equations (4), when all of the prices are set equal to *one*.

It is not at all correct, for a general selection of the individual demand functions x_1, x_2, x_3, that the market demand functions f_1, f_2, f_3 (4) must have a *unique* equilibrium point at (1, 1, 1). There are a number of very simple conditions, however, which imply that this equilibrium is unique. We shall give only one such set of conditions.

LEMMA 2. *Let*

$$A = \partial f_1/\partial p_3 < 0, \text{ and } B = \partial f_1/\partial p_2 > 0,$$

everywhere in the orthant $p_1 \geq 0, p_2 \geq 0, p_3 \geq 0$. *Then the equilibrium point* (1, 1, 1) *is unique (aside from positive multiples).*

In order to demonstrate this lemma, let us first make the observation that

$$\begin{aligned} f_2(p_1, p_2, p_3) &= f_1(p_2, p_3, p_1), \\ f_3(p_1, p_2, p_3) &= f_2(p_2, p_3, p_1), \\ f_1(p_1, p_2, p_3) &= f_3(p_2, p_3, p_1). \end{aligned}$$

This implies that if (α, β, γ) is an equilibrium point, then so is (β, γ, α), and also (γ, α, β). Therefore, if we have an equilibrium point different from (1, 1, 1), then we may find an equilibrium point (α, β, γ), perhaps by permutation, with either

$$\alpha \geq \beta \geq \gamma \tag{5}$$

(at least one of the inqualities is strict), or else

$$\alpha \leq \beta \leq \gamma \tag{6}$$

(again the same remark).

Let us assume that the former holds. An argument similar to the one we are about to give works if instead of (5) we have (6).

We shall show that $f_3(\alpha, \beta, \gamma)$ must be different from zero. Now

$$f_3(\alpha, \beta, \gamma) = x_3(\alpha, \beta, \gamma) + x_3(\beta, \gamma, \alpha) + x_3(\gamma, \alpha, \beta) .$$

The assumptions of this lemma imply that as a function of α, f_3 is strictly increasing, and as a function of γ, f_3 is strictly decreasing. Since $\alpha \geq \beta \geq \gamma$ (with inequality somewhere), we have

$$f_3(\alpha, \beta, \gamma) > x_3(\beta, \beta, \beta) + x_3(\beta, \beta, \beta) + x_3(\beta, \beta, \beta) = 0 ,$$

by the Walras Law. This completes the proof of the lemma.

Subsequently, in this paper, we shall have occasion to consider specific examples of demand functions for which the conditions of Lemma 2 do not hold for all prices in the orthant $p_1 \geq 0$, $p_2 \geq 0$, $p_3 \geq 0$, but rather for certain subsets. We shall then have recourse to the following lemma, which involves a type of subset convenient for us. The lemma is demonstrated in the same manner as Lemma 2. It should be noted that the operations involved in the proof of Lemma 2 do not take us out of the set described in Lemma 3.

LEMMA 3. *Let the conditions of Lemma 2 hold for the set of prices* $(p_1/p_3 > \varepsilon, p_2/p_1 > \varepsilon, p_3/p_2 > \varepsilon)$, *where* ε *is a small positive number. Then there are no equilibrium points in this set other than* (1, 1, 1) *and its multiples.*

We shall now give a condition on the demand functions which implies that the equilibrium point (1, 1, 1) is unstable.

LEMMA 4. *If* $C = \partial f_1/\partial p_1$ *is positive, at the point* (1, 1, 1), *then this equilibrium point is locally unstable in the following sense: there is a region (on the sphere* $\sum p_i^2 = 3$*) about the point* (1, 1, 1) *such that any point in this region (other than the equilibrium point) will move away from the point* (1, 1, 1).

The Jacobian of the functions f_1, f_2, f_3, at the point (1, 1, 1) is given by

$$\begin{pmatrix} C & A & B \\ B & C & A \\ A & B & C \end{pmatrix},$$

all evaluated at that particular point. Moreover, the functions f_1, f_2, f_3 are homogeneous of degree zero, and therefore the Euler relationship, at the point (1, 1, 1), tells us that

$$A + B + C = 0 . \tag{7}$$

The Jacobian, therefore, has one characteristic root equal to zero. The other characteristic roots may be shown to have positive real parts, because of the assumption $C > 0$. It is then possible to apply standard theorems of differential equations (see [4], chap. 13), in order to deduce local instability. However, the various properties of demand functions permit a simple independent proof of this fact, which we shall reproduce here.

Let

$$V(p_1, p_2, p_3) = \frac{1}{2}\sum (p_i - 1)^2 ,$$

so that

$$\frac{dV}{dt} = \sum (p_i - 1)\frac{dp_i}{dt} = -\sum f_i .$$

We shall show that $\sum f_i < 0$ in a small region on the sphere $\sum p_i^2 = 3$ about (1, 1, 1), (except for this point itself), and this will demonstrate the lemma. Of course at (1, 1, 1), $\sum f_i = 0$. Using the Taylor series expansion, we obtain

$$\sum f_i = \sum_j (p_j - 1)\sum_i \frac{\partial f_i}{\partial p_j} + \frac{1}{2}\sum_j \sum_k (p_j - 1)(p_k - 1)\sum_i \frac{\partial^2 f_i}{\partial p_j \partial p_k} + 0(V^{3/2}) , \tag{8}$$

where all of the partial derivatives are evaluated at the point (1, 1, 1). Let us simplify some of the terms. The Walras Law reads $\sum p_i f_i \equiv 0$, so that differentiating with respect to p_j, we obtain

$$f_j + \sum_i p_i \frac{\partial f_i}{\partial p_j} \equiv 0 . \tag{9}$$

At (1, 1, 1) this implies $\sum_i \partial f_i / \partial p_j = 0$, and therefore the linear terms of the Taylor series vanish. Differentiating (9) once more with respect to p_k, we obtain, at the point (1, 1, 1),

$$\sum_i \frac{\partial^2 f_i}{\partial p_j \partial p_k} = -\left(\frac{\partial f_j}{\partial p_k} + \frac{\partial f_k}{\partial p_j}\right). \tag{10}$$

From (7), (8), and (10) we see that

$$\sum_i f_i = -\frac{1}{2}\sum_{j,k}(p_j - 1)(p_k - 1)c_{jk} + 0(V^{3/2}),$$

where

$$(c_{jk}) = \begin{pmatrix} 2C, & -C, & -C \\ -C, & 2C, & -C \\ -C, & -C, & 2C \end{pmatrix}.$$

Therefore

$$\sum f_i = -\frac{C}{2}\{(p_1 - p_2)^2 + (p_2 - p_3)^2 + (p_3 - p_1)^2\} + 0(V^{3/2}).$$

But I claim that on the surface $\sum p_i^2 = 3$, with all $p_i \geq 0$, we always have

$$(p_1 - p_2)^2 + (p_2 - p_3)^2 + (p_3 - p_1)^2 \geq \frac{3}{2}\sum (p_i - 1)^2. \tag{11}$$

Accepting this as correct for the moment, we see that

$$\sum f_i < -\frac{3}{2}CV + 0(V^{3/2}),$$

and therefore $\sum f_i$ is negative for V sufficiently small.

Equation (11) is demonstrated as follows. First of all, on this sphere

$$3 \geq \sum p_i.$$

Multiplying by $\sum p_i$ and expanding, we obtain

$$3\sum p_i \geq p_1^2 + p_2^2 + p_3^2 + 2(p_1p_3 + p_1p_2 + p_2p_3),$$

and by subtracting 9 from each side we obtain

$$\frac{3}{2}(-\sum p_i^2 + 2\sum p_i - 3) \geq -2(p_1^2 + p_2^2 + p_3^2) + 2(p_1p_3 + p_1p_2 + p_2p_3),$$

which is the same as (11). This finishes the proof of Lemma 4.

We shall shortly exhibit some excess demand functions, derivable from a utility function, which satisfy the hypothesis of both Lemmas 2 and 4. They will provide us with examples of a unique equilibrium point, which is unstable locally. Does this imply global instability? In a certain sense this becomes a matter of definition, the problem being that it is possible that one of the prices might become zero. In this event, some care should be taken, since the differential equa-

tions (2) no longer describe the motion of the prices. In our examples, we shall use special techniques to demonstrate that the paths actually stay away from the boundary of the positive orthant.

Now let us consider utility functions for the first consumer of the form

$$U(x_1, x_2, x_3) = -\left(\frac{\alpha_1^{a+1}}{x_1^a} + \frac{\alpha_2^{a+1}}{x_2^a} + \frac{\alpha_3^{a+1}}{x_3^a}\right),$$

where $(\alpha_1, \alpha_2, \alpha_3)$ is an arbitrary non-negative vector, and a is a positive constant. Let the initial endowment of the first consumer be represented by (I_1, I_2, I_3). As before, we permute the utility function cyclically for the second and third consumers, and also permute the initial holdings.

We shall show that instability arises whenever $a > 1$, $(\alpha_1, \alpha_2, \alpha_3)$ is close to $(b, 1, 0)$ with $b > (a+1)/(a-1)$, and (I_1, I_2, I_3) is close to $(1, 0, 0)$. (The specific meaning of "close" will be clarified.) This will give us examples with none of the disagreeable features of the simple examples in Section 2. Let us first examine in detail the case

$$(\alpha_1, \alpha_2, \alpha_3) = (b, 1, 0),$$
$$(I_1, I_2, I_3) = (1, 0, 0),$$

with

$$b > \frac{a+1}{a-1}.$$

A routine calculation shows us that the excess demand functions for the first individual are given by

$$x_1 = \frac{bp_1^{a/(1+a)}}{bp_1^{a/(1+a)} + p_2^{a/(1+a)}} - 1,$$

$$x_2 = \frac{1}{p_1^{1/(1+a)}} \frac{p_2}{bp_1^{a/(1+a)} + p_2^{a/(1+a)}}, \tag{12}$$

$$x_3 = 0.$$

The reader should notice that if $b = 1$, and $a \to \infty$, these excess demand functions tend to those of Section 2.

Let us show that the conditions of Lemma 2 are verified. We have

$$A = \frac{\partial x_1(p_1, p_2, p_3)}{\partial p_2} < 0,$$

$$B = \frac{\partial x_2(p_3, p_1, p_2)}{\partial p_3} > 0.$$

Therefore the system of market demand functions (4) has a unique equilibrium point given by (1, 1, 1). In order to show that it is unstable we compute

$$C = \frac{\partial f_1}{\partial p_1}$$

at (1, 1, 1), or

$$C = \frac{ab - (b+1) - a}{(b+1)^2(a+1)},$$

which is positive if $b > (a+1)/(a-1)$. It follows from Lemma 4 that for this relationship between a and b the unique equilibrium point is also locally unstable.

The next step will be to show that for this example the prices stay away from the boundary of the positive orthant. This will be a consequence of the following lemma.

LEMMA 5. *Let*

$$\varphi = \text{Min}\left(\frac{p_2}{p_1}, \frac{p_3}{p_2}, \frac{p_1}{p_3}\right).$$

There is a small positive constant K, such that if

$$\varphi < K, \text{ then } \frac{d\varphi}{dt} > 0.$$

Let us assume that the minimizing value above is $p_1/p_3 = c$, say. The other cases are handled by a cyclic permutation. We want to show that $d(p_1/p_3)/dt > 0$ or $p_3 f_1 - p_1 f_3 > 0$, if c is sufficiently small. If f_1 and f_3 are computed according to (4), we see that $p_3 f_1 - p_1 f_3$ has the sign of

$$(13) \qquad \frac{\left(\frac{p_3}{p_1}\right)^a + 1}{b\left(\frac{p_3}{p_1}\right)^{a/(1+a)} + 1} - \frac{\frac{p_3}{p_1}}{b\left(\frac{p_1}{p_2}\right)^{a/(1+a)} + 1} - \frac{\frac{p_3}{p_2}}{b\left(\frac{p_2}{p_3}\right)^{a/(1+a)} + 1}.$$

The first term of (13) is equal to

$$\frac{\left(\frac{1}{c}\right)^a + 1}{b\left(\frac{1}{c}\right)^{a/(1+a)} + 1},$$

and since $p_2/p_3 \leq 1/c$, we have

$$\frac{-\frac{p_2}{p_3}}{b\left(\frac{p_2}{p_3}\right)^{a/(1+a)}+1} \geq \frac{-\frac{1}{c}}{b\left(\frac{1}{c}\right)^{a/(1+a)}+1}.$$

Also we have

$$\frac{-\frac{p_3}{p_1}}{b\left(\frac{p_1}{p_3}\right)^{a/(1+a)}+1} > -\frac{1}{c}.$$

Adding these terms together we see that (13) will be positive if

$$\left(\frac{1}{c}\right)^3 + 1 - \frac{2}{c} - b\left(\frac{1}{c}\right)^{1+a/(1+a)} > 0,$$

and this is certainly correct for small c. This demonstrates Lemma 5.

These observations, taken together, show us that the system of demand functions defined by (12) gives rise to global instability of the price adjustment mechanism.

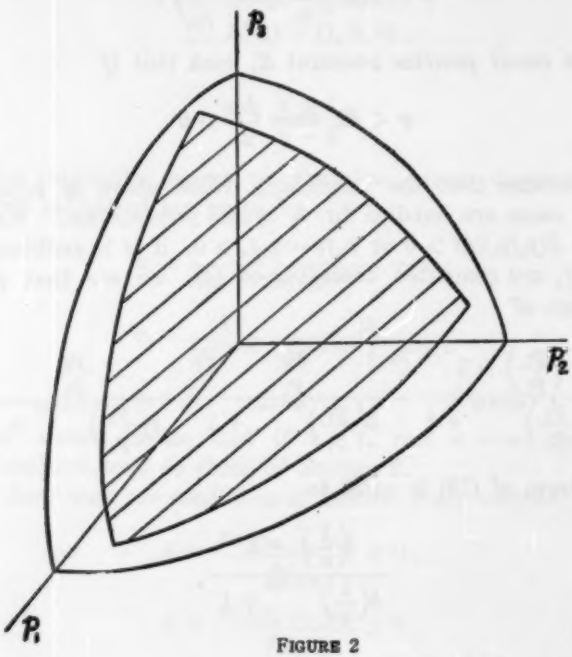

FIGURE 2

One more remark is in order. I claim that if we take an excess demand which is close to (12) on a large proper subset of the positive orthant (and which has derivatives which are close) and form the market demand functions using the above method of cyclic permutation, then this will also give rise to global instability. (Such an example may be obtained by taking $(\alpha_1, \alpha_2, \alpha_3)$ close to $(b, 1, 0)$ and (I_1, I_2, I_3) close to $(1, 0, 0)$, or in other ways.)

In order to see this let us interpret Lemma 5 geometrically.

Figure 2 represents the surface of the sphere $\sum p_i^2 = 3$, and the shaded region the set of points with

$$\text{Min}\left(\frac{p_2}{p_1}, \frac{p_3}{p_2}, \frac{p_1}{p_3}\right) \geq K,$$

for a small value of K. Lemma 5 tells us that a path which begins in this region, with K sufficiently small, will never leave the region. This is true because on the boundary of this region expressions such as $p_3 f_1 - p_1 f_3$ are strictly positive. But this later fact will also be true for market demand functions based on individual demand functions close to (12).

It will also be true that inside this region there will be no equilibrium points other than $(1, 1, 1)$. For if the new demand functions and their derivatives are close to the corresponding quantities for (12) inside the shaded region, then A for the new demand functions will be negative, and since it is homogeneous of degree -1 (being a derivative), A will be everywhere negative in the region defined by

$$\frac{p_2}{p_1} > K, \quad \frac{p_3}{p_2} > K, \quad \frac{p_1}{p_3} > K.$$

A similar statement with a reversed inequality holds for B, and therefore Lemma 3 may be applied to show that there are no equilibrium points other than $(1, 1, 1)$ in the shaded region.

A similar argument convinces us of the local instability of the equilibrium point $(1, 1, 1)$. It should be remarked that even though we have said nothing about possible equilibrium points outside the shaded region for the new demand functions, a path beginning inside the shaded region will not leave this region.

Stanford University and
Cowles Foundation at Yale University, U.S.A.

REFERENCES

[1] ARROW, KENNETH J., H. D. BLOCK, AND LEONID HURWICZ, "On the Stability of the Competitive Equilibrium, II," *Econometrica*, Vol. 27 (1959), pp. 82–109.

[2] ARROW, KENNETH J., AND GERARD DEBREU, "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, Vol. 22 (1954), pp. 256–290.

[3] ARROW, KENNETH J., AND LEONID HURWICZ, "On the Stability of the Competitive Equilibrium, I," *Econometrica*, Vol. 26 (1958), pp. 522–552.

[4] CODDINGTON, EARL A., AND NORMAN LEVINSON, *Theory of Ordinary Differential Equations* (New York: McGraw-Hill, 1955).

[5] HICKS, JOHN R., *Value and Capital* (Oxford: Oxford University Press, 1939).

[6] METZLER, LLOYD, "Stability of Multiple Markets: The Hicks Conditions," *Econometrica*, Vol. 13 (1945), pp. 277–292.

[7] SAMUELSON, PAUL A., "The Stability of Equilibrium: Comparative Statics and Dynamics," *Econometrica*, Vol. 9 (1941), pp. 97–120.

[8] UZAWA, HIROFUMI, "Walras Tâtonnement in the Theory of Exchange," Technical Report No. 69, Project on Efficiency in Decision Making in Economic Systems, Stanford University, March 1959.

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3, September, 1960

SIMULTANEOUS LEAST SQUARES: A DISTRIBUTION FREE METHOD OF EQUATION SYSTEM STRUCTURE ESTIMATION*

By T. M. Brown

SUMMARY

A METHOD of structure estimation is here developed which separates the regular from the random component of structure by minimizing the vector distance between the observed data and the corresponding values computed from the systematic structure of a complete model treated as a whole. The method is, accordingly, the simultaneous equation counterpart of single equation least squares. It too is distribution free.

The sample error variance-covariance matrix is constructed. Its elements are zero for an infinitely large sample. This, plus the hypothesis that only true structure could produce the above minimum vector distance, suggests consistency. The method is tested on a model with known structure, which generated its own data and moments. The estimates are equal to the true values.

The method is also tested on a model with unknown structure, using actual data. The results are compared with estimates by other well-known methods. Further research on the relative efficiency of estimation methods is suggested.

1. INTRODUCTION

There are three uncomfortable assumptions involved in the method of full information maximum likelihood (FML) for the estimation of economic structure. One is that the disturbance vector follows a multivariate *normal* distribution. A second is that a *large sample* of basic data is available for estimation purposes. And the third is that the best basic principle on which to base the estimation procedure is that of *maximum likelihood.*

With regard to the first assumption, while the distributions are probably bell-shaped and may tend asymptotically to normality, since each disturbance tends to combine a large number of lesser random influences, the normality assumption may not be closely met at all in

* Manuscript received February 26, 1960.

small samples. But small samples are the rule with economic time series appropriate for structure estimation of complete models. The passage of time may not correct this, for structure evolution may force the gradual rejection of data from periods too far in the past. This tends to invalidate the first two assumptions.

With regard to the third assumption, while optimal properties have been proved for the maximum likelihood (ML) approach to structure estimation, optimal properties have also been proved for the least squares (LS) approach. This produces a conflict of loyalties, a dichotomy. To demonstrate this point, let us assume that the disturbance of a single linear relation satisfying conditions of the Markoff theorem on least squares (David and Neyman [6]) is normally distributed. Then the ML and LS approaches lead to identical results, which are best linear unbiased estimates. There is no conflict. But let us next assume that the disturbance follows a non-normal, say a Gamma, distribution. Then the ML and LS estimation equations are quite different. Yet the LS estimates retain the property of being the best linear unbiased estimates, and in addition are much easier to compute. In this case which method, ML or LS, will produce better estimates, if we ignore computation costs?

In the case where the parent distribution of the random variable is not known, there is no conflict. LS must be used. It can be used with equanimity because it produces best unbiased estimates of structure. In addition, its sampling error formulas are appropriate for samples of any size.

The LS method appeals to the principle that minimizing the vector distance between the observed data and the systematic component in its explanation produces optimal structure estimates. Intuitively, one might assume that this is a necessary condition on the true structural values, when the universe of data is available. Some regular component added to the random component because of imperfect estimation of structure would increase this vector distance[1]. Conversely, it would be impossible to decrease this vector distance between true systematic and observed data by shifting some of the random component into the systematic.

These same concepts can be readily carried over to the complete system of simultaneous equations comprising an economic model. Let Y, Z, and u be vectors of endogenous, predetermined, and disturbance

[1] Let μ_i be the true random component and let r_i be part of the regular structure which is added to μ_i because of faulty structure estimation. $E(\mu_i+r_i)^2=E\mu_i^2+2E\mu_i r_i+Er_i^2$. Let $d^2=\sum(\mu^2+r^2)$, $\delta^2=\sum\mu^2$; $d^2>\delta^2$. Note that $E\mu_i r_i=0$ since r_i is part of regular structure.

variables respectively, in a complete economic model. Let $(YZ) = X$ and $x = X - \bar{X}$. Let B, C, and S be the matrices of unknown structural parameters in the model which are to be estimated, such that

$$By' + Cz' = u'; \tag{1}$$

$$E[u'u] = \sum; \quad \frac{\sum^{T} u_t' u_t}{T} = S; \quad Ax' = u'; \quad A = [B\,C]; \quad B \text{ is square}.$$

This complete system generates the endogenous variables from the structure, the predetermined data, and the disturbances through the process

$$y' = -B^{-1}Cz' + B^{-1}u' = Fz' + u_s'. \tag{2}$$

This is the simultaneous solution of (1) and produces the final values of y after all the interactions among the y have been taken into account.

The complete explanation of y separates into a regular component y_c, computed from the regular part of the structure, plus a random component u_s arising from the random disturbances in the system. Thus

$$y_c' = -B^{-1}Cz' = Fz', \tag{3}$$

$$u_s' = B^{-1}u', \tag{4}$$

$$y - y_c = u_s. \tag{5}$$

Equations (2) and (3) may be referred to as the *structural reduced form* of the system, the reduced form based on the structural parameters. Since y and z represent observed data, u_s represents the difference between the observed data and the component explained by the regular structure for the economic model treated as a whole. Parallel with the least squares principle for a single equation the following questions are pertinent: (a) What values of B and C will cause the vector distance of $u_{st}(t = 1, 2, \cdots, T)$ to be a minimum? (b) Will such values of B and C have optimal properties, as in the single equation case, e.g., will they be consistent? Question (a) can be answered by a well defined mathematical process which produces unique results[2]. With regard to (b), using the same kind of intuitive reason-

[2] The minimization of a vector distance in the space of endogenous variables suggests, in analogy with the theory of orthogonal regression, a lack of uniqueness under transformations of the endogenous variables. For any given units of measurement the results are unique. It can be shown that the results are invariant under orthogonal transformation of endogenous variables but not necessarily under any transformations in finite samples. In just identified models the results are invariant under linear transformations generally. I am indebted to Professor H. Theil for pointing out the problems of invariance under transformation.

ing as that above for the single equation case, we would expect to find consistency.[3]

Because of its parallelism in concept to single equation LS, the method suggested can be referred to as *Simultaneous Least Squares* (SLS). It takes fully into account the simultaneous interactions of all of the endogenous variables in the complete system. It obeys "Markoff conditions" within the multi-equation system, since it works with the y's as linear functions of z's and u_s's only, where the z's and u_s's are statistically independent.

The vector distance which is to be minimized is d, where

$$d^2 = \sum_{t=1}^{T} \sum_{i=1}^{g} u_{it}^2 . \tag{6}$$

There are g endogenous variables and equations in the system, k predetermined variables, and T time periods, for each of which y_t, z_t are observed. We also have

$$d^2 = D = \operatorname{tr} \sum_{t=1}^{T} u_{it}' u_{jt} = \operatorname{tr} TS_s = \operatorname{tr} S_s^* . \tag{7}$$

Thus the SLS method involves minimizing the trace of the variance-covariance matrix of the *structural* reduced form disturbances.

This reveals that SLS is closely related to FML, or to Quasi Full Maximum Likelihood (QFML), in the case where the FML procedure is used when the distribution of the u's is unknown. For QFML involves finding the values of B and C (in the reduced model with all identities eliminated) which maximize the function

$$L(a) = \log \frac{|B|}{|S|^{1/2}}, \tag{8}$$

([11] pp. 212-3; also [3]). These values will also maximize $|B|^2/|S|$, and hence will *minimize* $|B|^{-1}|S||B|^{-1}$. By the multiplication theorem for determinants $|B||B^{-1}| = 1$, and $|B^{-1}||S||B^{-1'}| = |B^{-1}SB^{-1'}| = \det S_s$. Thus QFML minimizes $\det S_s$ while SLS minimizes $\operatorname{tr} S_s$.

Since $\det S_s$ includes all of the covariances as well as the variances in S_s, while $\operatorname{tr} S_s$ includes only variances, it might be thought that this would in itself make QFML a more comprehensive method, using more information in the estimation procedure. But note that $\operatorname{tr} S_s = \operatorname{tr} B^{-1}SB^{-1'}$ involves all of the covariances as well as the variances of S_s and hence includes all of the covariances of the *primary* disturbance

[3] Consistency is the property that the estimates converge with probability 1 to the true values, as the sample size becomes infinite. A formal proof is given in the accompanying note in this issue by M. Nakamura.

vector u. Also min det S_r permits large variances to be offset by large *negative* covariances occurring in the right positions. Min tr S_r contains only the (positive) variances in S_r and, in addition, appears to have more meaning than min det S_r in relation to the estimation problem, as discussed above.

Which of the two methods is in fact superior can only be decided by further research. Questions of consistency (QFML was proved to be consistent in [11] Chap. II), sampling errors, and computational cost must be studied and weighed.

2. THE SLS ESTIMATION PROCEDURE

In the QFML method (cf [3], [4], [11]) it is necessary to eliminate all identities contained in (1). For SLS this is not necessary, but it seems advisable to do so anyhow, to reduce the size of the matrices in the computations. The reduced model with identities eliminated can be represented by

$$A_r x_r' = u_r', \quad \text{or} \quad B_r y_r' + C_r z_r' = u_r' . \tag{9}$$

Assume that there are g_r elements in y_r and k_r elements in z_r. Let the unrestricted or free structural parameters in A be $(a) = (a_1 \cdots, a_n)$ and let A also contain the known or restricted parameters $1, \bar{a}_p, \bar{a}_q$, and $a_s = ka_h$, where a_h is contained in vector (a). $\bar{a}_p, \bar{a}_q$, and k are assumed known in advance, possibly from cross-section and engineering measurements. In the estimation procedure they are always used with their known values. The individual elements a_{ri} of A_r are functions of the parameters of A.

$$a_{ri} = g_i(a_1, \cdots, a_n; 1, \bar{a}_p, \bar{a}_q, k) = f_i(a); \quad (i = 1, \cdots, m) . \tag{10}$$

From (9), we have

$$y_r' + B_r^{-1}C_r z_r' = B_r^{-1}u_r' = u_r' .$$

Then

$$D_r = \operatorname{tr} S_{rr} = \operatorname{tr} \sum^{T} (y_r' - F_r z_r')(y_r - z_r F_r') = D_r(a_r) = D(a) . \tag{12}$$

We now drop the subscript r for the sake of economy, but it is understood throughout. We also adopt the notation for moments.

$$m_{x_i x_j} = \sum^{T} x_i x_j; \quad x_i = X_i - \bar{X}_i; \quad [m_{x_i x_j}] = M_{zz}; \quad \{i, j = 1, 2, \cdots, (g + k)\} . \tag{13}$$

Then (12) becomes

$$D(a) = \operatorname{tr} (M_{yy} - M_{yz}F' - FM_{zy} + FM_{zz}F') . \tag{14}$$

This is a curved surface in the space of $(a_1, \cdots, a_n; D)$ where $a_1, \cdots, a_n$ are the unrestricted free parameters of A in (1).

Our problem now is to find the values $(\bar{a})$ which bring $D(a)$ to its minimum value $\bar{D}$. An iterative process is used. An initial vector $(a)_0$ somewhere near $(\bar{a})$ is obtained from some simpler single equation method. The direction of steepest descent from $(a)_0$ is given by the vector $-(\partial D/\partial a)_0$.

To obtain fast convergence (or even any convergence) this vector direction must be corrected from steepest to shortest descent. This is done by bringing it closer to a geodesic line through $(a)_0$ and leading toward $(\bar{a})$. This correction is achieved with the matrix $[\partial^2 D/\partial a' \partial a]_0^{-1} = D_2$ (Cf. Brown [3], Chernoff and Divinsky [4], and Crockett and Chernoff [5]). The iteration formula is

$$(15) \qquad (a)_1 = (a)_0 + \Delta(a)_0; \quad \Delta(a)_0 = -h\left(\frac{\partial D}{\partial a}\right)_0 \left[\frac{\partial^2 D}{\partial a' \partial a}\right]_0^{-1}.$$

As in the FML case, it is to be hoped that some simple component matrix, say P of D_2, will be found to act as a substitute for D_2, providing good convergence at reduced cost.

2.1. *Formula for $\partial D/\partial a$.* Formula (14) suggests that we first find $dD = (\partial D/\partial f)df'$, where $f = \operatorname{vec} F$. From (14)

$$(16) \qquad \begin{aligned} dD &= \operatorname{tr}\{-(dF)M_{yy} - M_{yy}dF' + FM_{xx}dF' + dFM_{xx}F'\} \\ &= 2 \operatorname{vec}(FM_{xx} - M_{yx}) \operatorname{vec}' dF, \end{aligned}$$

$$(17) \qquad \frac{\partial D}{\partial f} = 2 \operatorname{vec}(FM_{xx} - M_{yx}).$$

Then

$$(18) \qquad \frac{\partial D}{\partial a} = \frac{\partial D}{\partial f}\left[\frac{\partial f'}{\partial a}\right].$$

As $\bar{D}$ is approached, $\partial D/\partial a \to 0$. Note that from (17), (3), and (5), using z as a vector of instrumental variables, we obtain,

$$(19) \qquad \frac{\partial D}{\partial f} = 2 \operatorname{vec} \sum^{T} u_t' z.$$

This approaches zero for a large sample, but need not be zero for a small sample. For a small sample SLS does not force $\partial D/\partial f$ to be zero, and $\partial D/\partial a$ is zero by virtue of the product of the matrices in (18). But note that in the case of the *statistical reduced form* the LS regression of each y on all z's, $FM_{xx} - M_{yx}$, is equal to zero for samples of all sizes. This source of small sample error in all single equation LS estimates may not be present in SLS.

2.2. *Computation Formulas for* $[\partial f/\partial a']$.

$$F_r = -B_r^{-1}C_r , \tag{20}$$

$$dF_r = -B_r^{-1}dC_r - B_r^{-1}dB_rB_r^{-1}C_r , \tag{21}$$

$$\frac{\partial F_r}{\partial a_i} = -B_r^{-1}\left[\frac{\partial C_r}{\partial a_i}\right] - B_r^{-1}\left[\frac{\partial B_r}{\partial a_i}\right]F_r , \tag{22}$$

where a_i is an element of (a), not an element of (a_r).

$\partial C_r/\partial a_i$ and $\partial B_r/\partial a_i$ are easily computed from the structure of the reduced model. Numerical values of B_r^{-1} and F_r are readily available with each iteration. With these $\partial f/\partial a$ can be computed for each iteration. When working with a specific model, however, much time will be saved by applying (22) to the model structure and working out the algebraic formulas for each $\partial f_{hk}/\partial a_i$.

2.3. *Computation Formulas for* D_2. The direct differential approach to this matrix is tedious and impractical for large systems. However, a partial differentiation approach produces matrix formulas which can handle systems of any size.

Suppose $x = x(u, v)$, $y = y(u, v)$, then

$$\frac{\partial}{\partial u}(xy) = x\left(\frac{\partial y}{\partial u}\right) + \left(\frac{\partial x}{\partial u}\right)y . \tag{23}$$

The case is similar for $\partial(xy)/\partial v$.

$$dD = \left(\frac{\partial D}{\partial f}\right)df';\ df' = \left[\frac{\partial f'}{\partial a}\right]da';\ dD = \frac{\partial D}{\partial f}\left[\frac{\partial f'}{\partial a}\right]da' , \tag{24}$$

$$\frac{\partial^2 D}{\partial a' \partial a_i} = \frac{\partial}{\partial a_i}\left\{\left(\frac{\partial D}{\partial f}\right)\left[\frac{\partial f'}{\partial a}\right]\right\} . \tag{25}$$

Applying (23) to (25) we have

$$\begin{aligned}
\frac{\partial}{\partial a_1}\left\{\left(\frac{\partial D}{\partial f}\right)\left[\frac{\partial f'}{\partial a}\right]\right\} &= \frac{\partial D}{\partial f}\,\frac{\partial^2 f'}{\partial a_1 \partial a} + \left\{\frac{\partial}{\partial a_1}\left(\frac{\partial D}{\partial f}\right)\right\}\left[\frac{\partial f'}{\partial a}\right] \\
&\vdots \qquad\qquad \vdots \qquad\qquad \vdots \\
\frac{\partial}{\partial a_n}\left\{\left(\frac{\partial D}{\partial f}\right)\left[\frac{\partial f'}{\partial a}\right]\right\} &= \frac{\partial D}{\partial f}\,\frac{\partial^2 f'}{\partial a_n \partial a} + \left\{\frac{\partial}{\partial a_n}\left(\frac{\partial D}{\partial f}\right)\right\}\left[\frac{\partial f'}{\partial a}\right] ,
\end{aligned} \tag{26}$$

$$\begin{aligned}
d\left(\frac{\partial D}{\partial f'}\right) &= \frac{\partial^2 D}{\partial f' \partial f_1}df_1 + \cdots + \frac{\partial^2 D}{\partial f' \partial f_n}df_n \\
&= \left[\frac{\partial^2 D}{\partial f' \partial f}\right]\left[\frac{\partial f'}{\partial a}\right]da' .
\end{aligned} \tag{27}$$

Hence,

$$\frac{\partial}{\partial a}\left(\frac{\partial D}{\partial f'}\right) = \left[\frac{\partial^2 D}{\partial f' \partial f}\right]\left[\frac{\partial f'}{\partial a}\right]. \tag{28}$$

From (25), (26), and (28),

$$\left[\frac{\partial^2 D}{\partial a' \partial a}\right] = \left(\frac{\partial D}{\partial f}\right)\left|\left[\frac{\partial^2 f'}{\partial a' \partial a}\right]\right| + \left[\frac{\partial f}{\partial a'}\right]\left[\frac{\partial^2 D}{\partial f' \partial f}\right]\left[\frac{\partial f'}{\partial a}\right] = Q + P. \tag{29}$$

In (29), $|[\partial^2 f/\partial a' \partial a]|$ is a three dimensional array with gk rows (of the order of f), n columns (of the order of a), and n *layers* (of the order a). It is essentially a three-way matrix and could be designated as a *tri-mat*. The product of the vector $\partial D/\partial f$ and the tri-mat is a matrix Q of order n by n. It is computed in row by row sequence as follows:

$$\begin{aligned} \text{Row 1 of } Q &= \frac{\partial D}{\partial f}\left[\frac{\partial^2 f'}{\partial a_1 \partial a}\right] \\ &\vdots \qquad\qquad \vdots \\ \text{Row } n \text{ of } Q &= \frac{\partial D}{\partial f}\left[\frac{\partial^2 f'}{\partial a_n \partial a}\right]. \end{aligned} \tag{30}$$

The unknown portions of (29) so far are $\partial^2 f'/\partial a' \partial a$ and $\partial^2 D/\partial f' \partial f$. From (21) we know that $\partial f/\partial a$ consists of elements like $b^{vj} f_{kl}$ from B_r^{-1} and F. From these terms we build up the elements of $\partial^2 f/\partial a' \partial a$, using formula (23). Terms like $\partial b^{vj}/\partial a_m$ are computed from

$$dB^{-1} = -B^{-1} dB B^{-1}, \quad \frac{\partial B^{-1}}{\partial a_m} = -B^{-1}\left[\frac{\partial B}{\partial a_m}\right]B^{-1}, \tag{31}$$

$\partial^2 D/\partial f' \partial f$ being easily derived from (17).

$$\frac{\partial D}{\partial f} = 2 \operatorname{vec} F I_g \otimes M_{ss} - 2 \operatorname{vec} M_{gs}, \tag{32}$$

$$\frac{\partial^2 D}{\partial f' \partial f} = 2 I_g \otimes M_{ss}. \tag{33}$$

2.4. *P as a Substitute for D_2 for Iterations.* The P component of D_2 has all of its parts readily available and hence suggests itself as a convenient substitute for D_2 in the iterations. In the tests on Models X and Y below it proved to be satisfactory and gave rapid convergence.

3. THE SAMPLE ERROR VARIANCE-COVARIANCE MATRIX $S(a)$ OF THE SLS ESTIMATES

This section is sketched very briefly, to conserve space. Let E be the expectations operator, and d_s be the sampling differential operator,

as we assume theoretical repeated samples of Y and an estimate (a) corresponding to each sample. Z data are assumed constant in the repeated samples.

$$d_s a = a - E(a)\,. \tag{34}$$

Consider the change in $\bar{D}$ as we change from sample to sample. Omitting the bar and subscript s

$$dD = \frac{\partial D}{\partial a_1}da_1 + \cdots + \frac{\partial D}{\partial a_n}da_n + \frac{\partial D}{\partial y_{11}}dy_{11} + \cdots + \frac{\partial D}{\partial y_{gT}}dy_{gT}\,, \tag{35}$$

$$d\left(\frac{\partial D}{\partial a_i}\right) = \frac{\partial^2 D}{\partial a_i \partial a_1}da_1 + \cdots + \frac{\partial^2 D}{\partial a_i \partial a_n}da_n + \frac{\partial^2 D}{\partial a_i \partial y_{11}}dy_{11} + \cdots + \frac{\partial^2 D}{\partial a_i \partial y_{gT}}dy_{gT} = 0, \text{ as } D \to \bar{D},\ a \to \bar{a}\,; \tag{36}$$

$$\left[\frac{\partial^2 D}{\partial a' \partial a}\right]_T d_s a' + \left[\frac{\partial^2 D}{\partial a' \partial y_T}\right]_a d_s y'_T = 0 \quad (y_T = y_{11}, \cdots, y_{1T}, \cdots, y_{g1}, \cdots, y_{gT}), \tag{37}$$

$$E[da'da] = D_2^{-1}\left[\frac{\partial^2 D}{\partial a' \partial y_T}\right] E[dy'_T dy_T]\left[\frac{\partial^2 D}{\partial y'_T \partial a}\right] D_2^{-1}\,, \tag{38}$$

$$E[dy'_T dy_T] = \begin{bmatrix} \sum_{11}, \cdots, \sum_{1T} \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ \sum_{g1}, \cdots, \sum_{gT} \end{bmatrix}, \tag{39}$$

which reduces to $I_T \otimes \sum_g$ when we assume zero serial- and auto-correlation in the disturbances. From (17) and (18),

$$\frac{\partial D}{\partial a} = 2 \operatorname{vec}(FM_{xx} - M_{yx})\left[\frac{\partial f'}{\partial a}\right], \tag{40}$$

$$d_y\left(\frac{\partial D}{\partial a}\right) = -2 \operatorname{vec} d\begin{bmatrix} y_{11} \cdots y_{g1} \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ y_{1T} \cdots y_{gT} \end{bmatrix}' \begin{bmatrix} z_{11} \cdots z_{k1} \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ \cdot \quad\quad \cdot \\ z_{1T} \cdots z_{kT} \end{bmatrix} \left[\frac{\partial f'}{\partial a}\right], \tag{41}$$

$$\frac{\partial^2 D}{\partial y'_T \partial a} = -2 \operatorname{vec} \frac{\partial Y'_T}{\partial y'_T} Z_T \frac{\partial f'}{\partial a} = -2NI_g \otimes Z_T \frac{\partial f'}{\partial a}\,, \tag{42}$$

$$\begin{cases} \dfrac{\partial^2 D}{\partial y_{11} \partial a} = -2 \operatorname{vec} \dfrac{\partial Y'_T}{\partial y_{11}} Z_T \dfrac{\partial f'}{\partial a} \\ \quad\vdots \qquad\qquad \vdots \\ \dfrac{\partial^2 D}{\partial y_{gT} \partial a} = -2 \operatorname{vec} \dfrac{\partial Y'_T}{\partial y_{gT}} Z_T \dfrac{\partial f'}{\partial a}\,. \end{cases}$$

where N is $gT \times gT$. The small sample formula is

$$S(a) = 4D_2^{-1}\frac{\partial f}{\partial a'}I_g \otimes Z_T'N'I_T \otimes S_*NI_g \otimes Z_T\frac{\partial f'}{\partial a}D_2^{-1}. \tag{43}$$

For large samples we change the moment formula (13) to

$$M_{x_ix_j} = \frac{1}{T}\sum_1^T x_ix_j. \tag{44}$$

$S(a)$ now becomes dimensionally of the order $(4/T)V$, showing that $S(a) \to 0$ as $T \to \infty$.

4. GOODNESS OF FIT

All of the conventional goodness of fit statistics can be calculated, once the model has been estimated. A suggested selection is: $t_i = a_i/S(a_i)$; coefficients of variation, $V_i = \{S(u_i)/\bar{Y}_i\} \times 100$ and, $V_{si} = \{S(u_{si})/\bar{Y}_i\} \times 100$; coefficients of correlation, $R_i = [\{S^2(Y_i) - S^2(u_i)\}/S^2(Y_i)]^{1/2}$, and $R_{si} = [\{S^2(Y_i) - S^2(u_{si})\}/S^2(Y_i)]^{1/2}$; and, finally, von Neumann ratios for each u_i and u_{si}.

It is also of interest to consider the possibility of a single measure, corresponding to R, of the goodness of fit of the complete model treated as a whole. Recent work on this problem has been done by Hooper [8] and Theil [13]. In terms of the SLS estimates, the measure

$$R_s = \left(\frac{\text{tr } M_{yy} - \text{tr } S_*}{\text{tr } M_{yy}}\right)^{\frac{1}{2}} \tag{45}$$

gives a parallel to the R_{si} above. This formula, however, omits covariances. This omission can be remedied, if desired, by using determinants rather than traces in (45). We recall that trS_* already includes all of the covariances of S. An interesting variation which also includes covariances follows from Hooper's formulas and his suggestion on p. 254 of [8].

$$R_s = \left\{\frac{\text{tr}}{g}[I - M_{yy}^{-1}S_*]\right\}^{\frac{1}{2}}. \tag{46}$$

5. COMPUTATIONAL EXPERIMENTS

The SLS method was first tested on Model X, the true structure of which was known. This same model also had been used for testing in [3]. In this model α and Σ are given. Also assumed given is a fixed sample of Z_T. From these Z_T data, Model X generates its possible universe of Y data, from which population moments could be computed. From these moments experimental computations were made

to estimate the structure α and Σ, using the SLS method. If the derived estimates equal α and Σ, this would demonstrate the consistency of the SLS method. At the same time, the number of iterations to convergence and the general volume of computation could be appraised.

5.1. *Model X.*

$$
\begin{aligned}
&(1)\quad C = \underset{a_0}{.5} + \underset{a_1}{.8}W + \underset{a_2}{.4}\pi + u_1\,,\\
&(2)\quad W = \underset{b_0}{.4} + \underset{b_1}{.3}Y + \underset{b_2}{.2}Y_{-1} + u_2\,,\\
&(3)\quad W + \pi + T_g = Y,\\
&(4)\quad C + E = Y\,.
\end{aligned}
\tag{47}
$$

Endogenous Variables	*Predetermined Variables*
$C = Y_3$ = consumer expenditures	$T_g = Z_1$ = government net revenue
$W = Y_1$ = wage-salary bill	$E = Z_2$ = all non-consumer spending on newly produced final goods
$\pi = Y_2$ = non-wage income	
$Y = Y_4$ = total income, production, and expenditure	$Y_{-1} = Z_3$ = value of Y lagged one time period.

With identities eliminated, and variables converted to deviation from mean values, the reduced model becomes,

$$
B_r = \begin{bmatrix} (a_1 - 1) & (a_2 - 1) \\ (b_1 - 1) & b_1 \end{bmatrix};\quad C_r = \begin{bmatrix} -1 & 1 & 0 \\ b_1 & 0 & b_2 \end{bmatrix};\quad B_rC_r = A_r\,;
\tag{48}
$$

$$
x = (y_1, y_2, z_1, z_2, z_3) = (X - \bar{X});\quad u = (-u_1, -u_2);\quad A_rx' = u'\,.
$$

We assume the universe of u to produce a constant variance-covariance matrix.

$$
\sum = E(u'u) = \begin{bmatrix} .008 & .004 \\ .004 & .012 \end{bmatrix}.
\tag{49}
$$

To make this matrix more comparable to small sample formulas, we multiply it by a scalar $T = 25$.

$$
\sum{}^{*} = T\sum = \begin{bmatrix} .2 & .1 \\ .1 & .3 \end{bmatrix}.
\tag{50}
$$

Moment formulas are defined as follows:

$$
m^{p}_{z_i z_j} = E\,(X_i - \bar{X}_i)(X_j - \bar{X}_j)
\tag{51}
$$

is a population moment.

$$m_{z_i z_j} = \sum^{T} (X_i - \bar{X}_i)(X_j - \bar{X}_j)$$

is a small sample moment, with T observations in the sample. But for this experiment we also use $m_{z_i z_j} = T m^{p}_{z_i z_j}$, with T a scalar equal to 25. The z moments as derived from the Z_T data are assumed to be

$$M_{zz} = [m_{z_i z_j}] = TM^{p}_{zz} = \begin{bmatrix} 1.0 & .1 & .2 \\ .1 & 2.0 & .1 \\ .2 & .1 & 3.0 \end{bmatrix}. \tag{52}$$

The moments for the universe of detailed Y data which the above structure would generate from the given Z_T data can be computed directly from the formulas:

$$M_{yy} = TM^{p}_{yy} = B_T^{-1} \sum\nolimits^{*} B_T^{-\nu} + B_T^{-1} C_T M_{zz} C_T' B_T^{-\nu}, \tag{53a}$$

$$M_{yz} = -B_T^{-1} C_T M_{zz}. \tag{53b}$$

The subscript r is henceforth usually dropped from A, B, and C, but is understood. Also the lower triangle of all symmetric matrices is omitted. From formulas (53) we obtain

$$M_{yy} = \begin{matrix} y_1 \\ y_2 \end{matrix} \begin{bmatrix} 1.709375 & 2.159375 \\ & 6.703819 \end{bmatrix}, \tag{54a}$$

$$M_{yz} = \begin{matrix} y_1 \\ y_2 \end{matrix} \begin{matrix} \begin{bmatrix} -.137500 & 1.250000 & .762500 \\ -1.454167 & 2.750000 & -.420833 \end{bmatrix} \end{matrix} \begin{matrix} z_1 & z_2 & z_3 \end{matrix}. \tag{54b}$$

The free or unrestricted parameters of Model X are evident from (47). They are

$$\begin{aligned} (\mathrm{a}) &= (a_1, a_2, b_1, b_2), \\ (\alpha) &= (.8, .4, .3, .2). \end{aligned} \tag{55}$$

The same opening vector point from which iterations began in the FML experiment [3] is used in this experiment with SLS. Thus

$$(a)_0 = (.6, .3, .4, .1), \tag{56}$$

with percentage deviations from true values $(-25.0, -25.0, +33.3, -50.0)$. These deviations are of the order of $33\frac{1}{3}$ per cent.

The calculations for Model X were done by Mr. J. A. Aurèle Leduc, who took considerable interest in the work and made helpful suggestions. In the main, the calculations were carried out to ten significant figures to keep rounding errors from accumulating into the first four or five significant figures. The data were scaled so that ten significant figures usually meant nine decimal points. The numerical results

which follow are, however, shown only to six decimal points to save space. Matrix inversions were done by either adjugate or triangular methods, with the square root procedure applied in the case of symmetric matrices. Sum checks were used throughout for all matrix products and inversions.

i. *The First Iteration*:

$$(57) \qquad -B_{r0} = \begin{bmatrix} .400000 & .700000 \\ .600000 & -.400000 \end{bmatrix}; \; C_{r0} = \begin{bmatrix} -1 & 1 & 0 \\ .4 & 0 & .1 \end{bmatrix};$$

$$(58a) \qquad -B_{r0}^{-1} = \begin{bmatrix} .689655 & 1.206897 \\ 1.034483 & -.689655 \end{bmatrix};$$

$$(58b) \qquad -B_{r0}^{-1}C_{r0} = F_0 = \begin{bmatrix} -.206897 & .689655 & .120690 \\ -1.310345 & 1.034483 & -.068966 \end{bmatrix};$$

$$(59) \qquad H_0 = F_0 M_{ss} = \begin{bmatrix} -.113793 & 1.370690 & .389655 \\ -1.220690 & 1.931034 & -.365517 \end{bmatrix};$$

$$(60) \qquad -\frac{1}{2}\left(\frac{\partial D}{\partial f}\right)_0 = -\operatorname{vec} L_0 = \operatorname{vec}(M_{ys} - H_0)$$

$$= (-.023707, \; -.120690, \; .372845, \; -.233477, \; .818966, \; -.055316).$$

(a) *Formula for* $[\partial f/\partial a']$. Applying formula (22) to Model X gives

$$(61) \qquad \left[\frac{\partial f}{\partial a'}\right] = \left[\begin{array}{ccc|ccc} b^{11}F & & & & & \\ b^{12}\{(1+f_{11}+f_{21}), & (f_{12}+f_{22}), & (f_{13}+f_{23})\} & & & \\ 0 & 0 & b^{13} & & & \\ \hline & & & b^{21}F & & \\ & & & b^{22}\{(1+f_{11}+f_{21}), & (f_{12}+f_{22}), & (f_{13}+f_{23})\} \\ & & & 0 & 0 & b^{23} \end{array}\right],$$

where the b^{ij} are the elements of $-B_r^{-1}$ and the f_{ij} are elements of F.

$$(61a) \qquad \left[\frac{\partial f}{\partial a'}\right]_0$$

$$= \begin{bmatrix} -.142687 & .475624 & .083234 & -.214031 & .713436 & .124851 \\ -.903686 & .713436 & -.047562 & -1.355529 & 1.070155 & -.071344 \\ -.624257 & 2.080856 & .062426 & .356718 & -1.189061 & -.035672 \\ 0 & 0 & 1.206897 & 0 & 0 & -.689655 \end{bmatrix}.$$

(b) *Vector of Steepest Descent.*

$$-\frac{1}{2}\left(\frac{\partial D}{\partial a}\right)_0 = -\frac{1}{2}\left(\frac{\partial D}{\partial f}\right)\left[\frac{\partial f'}{\partial a}\right]_0 \tag{62}$$

$$= (.604358,\ \ 1.114437,\ \ -1.268175,\ \ .488134),$$

$$\frac{1}{2}\left[\frac{\partial^2 D}{\partial f'\partial f}\right] = I_3 \otimes M_{xx}\,. \tag{63}$$

(c) *The P Matrix.*

$$\frac{1}{2}P_0 = \left[\frac{\partial f}{\partial a'}\right]_0 \times \frac{1}{2}\left[\frac{\partial^2 D}{\partial f'\partial f}\right] \times \left[\frac{\partial f'}{\partial a}\right]_0$$

$$= \begin{bmatrix} 1.570316 & 2.380793 & .290401 & .046332 \\ & 5.599359 & .468710 & -.043462 \\ & & 11.689263 & .433085 \\ & & & 5.796671 \end{bmatrix}, \tag{64}$$

$$2P_0^{-1} = \begin{bmatrix} 1.796425 & -.762854 & -.013334 & -.010082 \\ & .503161 & -.001594 & .009989 \\ & & .086179 & -.006344 \\ & & & .173214 \end{bmatrix}. \tag{65}$$

(d) *Vector of Shortest Descent.*

$$-\left(\frac{\partial D}{\partial a}\right)_0 P_0^{-1} = d_0 = (.243125,\ \ .106601,\ \ -.122221,\ \ .092197), \tag{66}$$

$$(a)_1 = (a)_0 + h_0 d_0\,, \tag{67}$$

with h_0 selected at $0.5 = (.721563,\ \ .353301,\ \ .338890,\ \ .146098)$.

(e) *Distance Formula for Checking Progress of Iterations.* From (14) we have

$$D = \mathrm{tr}M_{yy} + \mathrm{vec}\,(FM_{xx} - M_{yx} - M_{yx})\,\mathrm{vec}'\,F$$

$$= \mathrm{tr}M_{yy} + \mathrm{vec}\,(L - M_{yx})\,\mathrm{vec}'\,F\,, \tag{68}$$

$$D_0 = 1.527716;\ \ D_0 = \bar{D} = 1.059028\,. \tag{69}$$

Progress in the iterations can also be checked from the changes in $\partial D/\partial f$ and $\partial D/\partial a$.

ii. *The Subsequent Iterations.* To save labor, the P_0 matrix was used for three iterations, and then a new matrix P_3 was computed for subsequent iterations. As the iterations proceeded, Mr. Leduc suggested that $\partial f/\partial a'$ also be kept constant for three iterations, to shorten the computations. In testing this proposal it was found that the temporarily constant $\partial f/\partial a'$ had the further advantage of giving faster convergence.

After six iterations, using only two P and two $\partial f/\partial a'$ matrices, the estimates $(a)_6$ had converged to the true values to at least six decimal points. The sequence of h values selected for the six iterations were (.5, .5, .75, 1.0, 1.0, 1.0).

Thus, we have, at the minimum, an empirical test that the SLS method, in minimizing the vector distance $(y_t - y_{ct})$, correctly separates and estimates the systematic and random structures, given the universe of data. Also, it was noted that at $\bar{D}$, $\frac{1}{2}(\partial\bar{D}/\partial f) = \text{tr}(FM_{zz} - M_{yz}) = 0$, a necessary condition on the true structure and population moments, (cf. (19)).

iii. *Truncated Simultaneous Least Squares (TSLS)*[5]. As in the case of FML [3], the cost of the SLS procedure can be considerably reduced by using a truncation procedure. This gain will, of course, be at the expense of a certain bias or loss in accuracy. The truncation tested for Models X and Y was to use a constant $[\partial f/\partial a]_0$ and P_0 matrix for four (or three) iterations and to terminate computations at this point. The results for four iterations on model X are shown in Table 1.

TABLE 1

TEST OF MODEL X STRUCTURE ESTIMATION BY TSLS

True Values		Estimates by TSLS	e-TSLS
α_1	.8	.802339	+0.3%
α_2	.4	.398539	−0.4
α_3	.3	.298840	−0.4
α_4	.2	.196402	−1.8
$\bar{D}$	1.059023	1.059137	+0.01%

The bias or error formula used in Table 1 is

$$e_i = \frac{a_i - \alpha_i}{\alpha_i} \times 100\% \,. \tag{70}$$

iv. *Test of Professor Theil's Two-Stage Least-Squares (TL) Method.* Before leaving Model X, it seemed like a good opportunity to test Professor Theil's interesting new single equation method [13]. It gave perfect estimates on Model X, verifying its consistency, for which there is, of course, theoretical proof. (See also Basmann [1]).

5.2. *Model Y.*

Model Y is a small model of the Canadian Economy, built from

[5] The initials TSLS are not here to be confused with those for "two-stage-least-squares," dealt with in the next section.

actual economic data which are now considerably out of date. The data used are for the years 1927–41 and 1946–52, comprising 22 observations. Though out of date, Model Y is useful for continued experimentation, since it is small and based on a sample of actual data. Also, since its data and structure remain constant, it permits comparison of estimation methods. It was used in this way in the experimentation on FML reported on in [3]. It will be recognized that Model Y corresponds, except for one variable, with the model reported on in [2]. This latter model covered only the years 1927–41 and 1946–49.

In this article Model Y is described in detail and its moments are given. This is done to enable readers to check the results shown here and in [3], and to perform any further testing which may occur to them.

MODEL Y

(71)

$$(1)\quad C = b_{12}Y_w + b_{13}Y_x + c_{11}C_{-1} + c_{12}L_{h,-1} + c_{10} + u_1\,,$$

$$(2)\quad Y_w = b_{24}(Y + S_1) + c_{23}(Y + S_1)_{-1} + c_{24}\frac{t}{10} + c_{20} + u_2\,,$$

$$(3)\quad Y = Y_w + Y_x\,,$$

$$(4)\quad Y + S_1 = C + S_0 + S_1\,.$$

All variables are measured in billions of 1935–39 constant dollars, except for t, which is measured in years (calendar year minus 1926)[6]. The variables are defined very briefly as follows:

(72)

$Y_1 = C$ = consumer spending on goods and services,

$Y_2 = Y_w$ = disposable wage-salary income,

$Y_3 = Y_x$ = disposable non-wage or property-enterprise income,

$Y_4 = Y + S_1 = GNP$,

$Z_1 = C_{-1} = C$ lagged one year,

$Z_2 = L_{h,-1}$ = liquid asset holdings of households at end of previous year,

$Z_3 = (Y + S_1)_{-1}$,

$Z_4 = t/10$,

$Z_5 = S_0 = Y - C$,

$Z_6 = S_1 = GNP - Y$.

The moment matrices, defined as in (13), are

[6] Mr. Lou Scott added his computational abilities to the work on Model Y, and his neatness, accuracy, and helpful co-operation are appreciated in enabling the work to be completed in the time available.

$$M_{yy} = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \begin{array}{c} \begin{array}{cccc} y_1 & y_2 & y_3 & y_4 \end{array} \\ \begin{bmatrix} 48.017285 & 38.077358 & 21.738878 & 77.633277 \\ & 30.494605 & 17.174202 & 62.191638 \\ & & 10.933463 & 35.894292 \\ & & & 128.449409 \end{bmatrix} \end{array},$$

$$M_{yz} = \begin{array}{c} \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{array} \begin{array}{c} \begin{array}{cccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 \end{array} \\ \begin{bmatrix} 45.083213 & 66.468980 & 74.176188 & 23.989559 & 11.798951 & 17.817041 \\ 35.676996 & 52.058769 & 59.058006 & 19.386505 & 9.591449 & 14.522831 \\ 20.163852 & 29.848450 & 33.532672 & 10.367895 & 6.368787 & 7.786627 \\ 72.553932 & 103.353066 & 120.752312 & 38.814432 & 20.452653 & 30.363479 \end{bmatrix} \end{array}, \tag{73}$$

$$M_{zz} = \begin{array}{c} \\ z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \\ z_6 \end{array} \begin{array}{c} \begin{array}{cccccc} z_1 & z_2 & z_3 & z_4 & z_5 & z_6 \end{array} \\ \begin{bmatrix} 43.196648 & 62.084527 & 69.908005 & 22.580436 & 10.757635 & 16.713084 \\ & 104.050177 & 101.694914 & 33.713850 & 15.438241 & 21.445845 \\ & & 116.458406 & 36.314400 & 18.414490 & 28.161634 \\ & & & 13.818636 & 5.764841 & 9.060032 \\ & & & & 4.161285 & 4.492417 \\ & & & & & 8.054021 \end{bmatrix} \end{array}.$$

The results of some of the various methods which have been used for estimating the structure of Model Y are listed in Table 2. Some methods not here listed are reported in [3]. SLS results are put in the lead position and error formula (70) is adapted by replacing α_i by a_{ii}, the SLS estimates. Either these or the QFML results are assumed to be the best available estimates of Model Y structure. In the computations on Model Y constant h values of 1 were used throughout, and six iterations gave $\partial D/\partial a = 0$ up to six decimal points. Iterations were started from $(a)_0$ = LS estimates, just as for the QFML estimates in [3]. P_0 and $[\partial f/\partial a']_0$ were used without change for iterations 1, 2, and 3, and then P_3 and $[\partial f/\partial a']_3$ were used for iterations 4, 5, and 6. In computing TSLS as a short-cut method, three iterations with the original P_0 and $[\partial f/\partial a']_0$ were found to give a closer approximation to the SLS values than four iterations. The results recorded in Table 2 suggest that if optimal single equation results were used for $(a)_0$—for example TL or LI (Limited Information)—TSLS would give results very close to those of SLS.

In this small sample case, SLS converged as rapidly and easily as for Model X. The vector $\partial D/\partial f$, however, did not reduce to zero, although most of its elements became small. On reflection we realize that (19) should not usually be zero for a small sample. Our method does not force this equality, and this may be one of its strengths.

TABLE 2

COMPARATIVE ESTIMATES OF MODEL *Y* STRUCTURE

Structural Parameters (a)	SLS-6 Iterations	QFML-10 Iterations	e-QFML	TSLS	e-TSLS	TL	e-TL	LI	e-LI
b_{12}	.678681	.680868	+0.3	.684290	+0.8	.682987	+ 0.6	.686677	+1.2
b_{13}	.064848	.063090	−2.7	.063965	−1.4	.087960	+35.6	.065352	+0.8
c_{11}	.351194	.350868	−0.1	.346262	−1.4	.335461	− 4.5	.341344	−2.8
c_{12}	.070813	.070369	−0.6	.071283	+0.7	.071707	+ 1.3	.072836	+2.9
b_{24}	.159513	.159723	+0.1	.161895	+1.5	.163828	+ 2.7	.161933	+1.5
c_{23}	.240246	.241150	+0.4	.236802	−1.4	.239683	− 0.2	.241373	+0.5
c_{24}	.032488	.032147	−1.0	.032782	+0.9	.031289	− 3.7	.031377	−3.4

6. CONCLUSIONS

The SLS method of structure estimation for econometric models has behaved well in the above tests. Its results compare favourably with QFML. It also seems to be easier than QFML to compute, both in individual iterations and in the number of iterations to convergence. Like QFML, it is a full information method.

SLS has appeal on theoretical grounds, for intuitive theory suggests that only the true systematic and random structures can produce the minimum vector distance between observed data and the systematic component of the explanation of these observed data. In following this approach, one need make no assumptions about the distribution of the random variables, or about the sample size. The method is accordingly both "distribution free" and "sample size free."

The present study suggests that it would now be useful for research to attempt to decide on the superiority of SLS or QFML. This would hinge partly on the relative sizes of $S(a)$ for the same model estimated by the two methods, and partly on any predisposition of either to produce small sample biases.

It may only be possible to compare these magnitudes when they are in numerical form, related to a specific numerical model. To obtain such numerical tests, synthetic models like Model *X* could be set up, and *samples* of data could be drawn from them by Monte Carlo methods. The synthetic model used could have the same specifications as some model in which the researcher had a particular interest. Estimates of both a and $S(a)$ could then be made by each of the two methods. It should be possible to work out the sample error formula for $S(a)$ of QFML, proceeding exactly as above for SLS, and merely replacing $D = \operatorname{tr} S$, with $D = \det S$,.

Any tendency of either method to a small sample bias and the size

of such bias as an empirical function of sample size could be estimated by drawing repeated Monte Carlo samples in groups of varying sample sizes from the model. For each sample the structure a could be estimated, preparing the way for estimates of the probability distribution of a for each sample size. Electronic computers make research of this kind a fascinating possibility. That method which reveals the smallest small sample biases and the smallest sampling error variances and covariances would win the first phase of the contest.

The second phase would consist of a comparison of computational cost. An efficiency function varying inversely with sample bias, sampling errors, and computational cost could then be set up to help choose the appropriate estimation method for a particular purpose. In making such efficiency comparisons it might be appropriate to include also the truncated and the single equation methods.

Royal Military College of Canada

REFERENCES

[1] BASMANN, R. L., "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," *Econometrica*, Vol. 25, No. 1 (Jan. 1957).

[2] BROWN, T. M., "Habit Persistence and Lags in Consumer Behaviour," *Econometrica*, Vol. 20, No. 3 (July 1952).

[3] BROWN, T. M., "Simplified Full Maximum Likelihood and Comparative Structural Estimates," *Econometrica*, Vol. 27, No. 4 (Oct. 1959).

[4] CHERNOFF, HERMAN AND NATHAN DIVINSKY, "The Computation of Maximum-Likelihood Estimates of Linear Structural Equations," Chap. 10 in Wm. C. Hood and Tjalling C. Koopmans, *Studies in Econometric Method*, Cowles Commission Monograph No. 14. New York: John Wiley and Sons, 1953.

[5] CROCKETT, JEAN BRONFENBRENNER AND HERMAN CHERNOFF, "Gradient Methods of Maximization," *Pacific Journal of Mathematics*, Vol. 5–1 (March 1955); also Cowles Commission Paper, New Series, No. 92.

[6] DAVID, F. N. AND J. NEYMAN, "Extension of the Markoff Theorem on Least Squares," *Statistical Research Memoirs*, Vol. II. London: Department of Statistics, University of London, University College, Dec. 1938, pp. 105–116.

[7] GEARY, R. C., "Determination of Linear Relations between Systematic Parts of Variables with Errors of Observation the Variances of Which are Unknown," *Econometrica*, Vol. 17, No. 1 (Jan. 1949).

[8] HOOPER, J. W., "Simultaneous Equations and Canonical Correlation Theory," *Econometrica*, Vol. 27, No. 2 (April 1959).

[9] KLEIN, LAWRENCE R., *A Textbook of Econometrics*. Evanston: Row, Peterson and Co., 1953.

[10] ———, "On the Interpretation of Theil's Method of Estimating Economic Relationships," *Metroeconomica*, Vol. 7, No. 3 (Dec. 1955).

[11] KOOPMANS, TJALLING C., ed., *Statistical Inference in Dynamic Economic Models*, Cowles Commission Monograph No. 10. New York: John Wiley and Sons, 1950.

[12] SARGAN, J. D., "The Estimation of Economic Relationships Using Instrumental Variables," *Econometrica*, Vol 26, No. 3 (July, 1958).

[13] THEIL, H., *Economic Forecasts and Policy*. Amsterdam: North-Holland Publishing Co., 1958.

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3. September, 1960

A NOTE ON THE CONSISTENCY OF SIMULTANEOUS LEAST SQUARES ESTIMATION*

BY MITSUGU NAKAMURA[1]

VARIOUS PROBLEMS connected with structural estimation through simultaneous least squares procedure have been studied by T. M. Brown in a preceding article in this issue[2]. In this paper he suggests that simultaneous least squares estimation has the property of consistency, as well as a lighter computational burden as compared with ordinary full-information quasi-maximum-likelihood estimation, with which his method may naturally be compared under fairly general conditions, without assuming any particular type of distribution. Brown shows the results of several experiments which produce favorable comparisons in plausible situations. Rigorous proof of consistency by this procedure, especially concerning sufficient conditions, is, however, open to question. In this note I shall try to present a proof for consistency of simultaneous least squares estimation starting from fairly general assumptions.

Our structural model is as follows:

$$By_t + \Gamma z_t = u_t \qquad (t = 1, 2, \cdots), \tag{1}$$

with assumptions:

$$|B| \neq 0, \tag{2}$$

$$E(u_t | z_t) = 0, \tag{3}$$

where y_t and u_t are column vectors with m components and z_t is a column vector with n components. Our model is complete, having m endogenous variables $(y_{1t}, y_{2t}, \cdots, y_{mt})$ and n exogenous variables $(z_{1t}, z_{2t}, \cdots, z_{nt})$.

Introducing new parameters Π and new random disturbances v_t, by definition, we have:

$$\Pi \equiv -B^{-1}\Gamma, \tag{4}$$

$$v_t \equiv B^{-1}u_t. \tag{5}$$

* Manuscript received, January 15, 1960.

[1] I am much indebted to Professor L. R. Klein and Professor T. W. Anderson for reading the manuscript and making useful suggestions. The writer alone, of course, is responsible for any errors which may remain.

[2] See pp. 173–192.

We now can obtain transformed relations, i.e., the so-called reduced forms:

$$y_t = \Pi z_t + v_t \,. \tag{6}$$

We assume the following conditions in the above structure:

i) Structural coefficient parameters, B, Γ, are identifiable from *a priori* information about B, Γ, and the square matrix B is non-singular.

ii) Conditional distributions of $(u_1, u_2, \cdots, u_N)$ given $(z_1, z_2, \cdots, z_N)$ are independent of the given value of $(z_1, z_2, \cdots, z_N)$.

iii) $u_t (t = 1, 2, \cdots)$ are random variables mutually independent with respect to different t and have bounded variances of their components.

iv) $|z_t| (t = 1, 2, \cdots)$ are bounded, say $|z| < K$, and are not zero.

v) $\underline{\lim} |z_t| \neq 0$.

vi) Define "Limit Directions" of $z_t (t = 1, 2, \cdots)$ as a set of limit vectors of $z_t/|z_t| (t = 1, 2, \cdots)$. Then we assume here that z_t has at least n mutually linearly independent limit vectors in "Limit Directions" of z_t.

We denote the sum of squares divided by the sample numbers as

$$\frac{1}{N}\sum_{t=1}^{N}(y_t - \Pi z_t)'(y_t - \Pi z_t) \equiv S_N^2(y_1, y_2, \cdots, y_N; z_1, \cdots, z_N; \Pi)$$

or more briefly

$$\equiv S_N^2(Y_N, \Pi) \,,$$

and the true coefficient structures as B_0, Γ_0, and Π_0 respectively, or $\Pi_0 \equiv -B_0^{-1}\Gamma_0$. We define simultaneous least squares estimates of Π from a sample of N, $\hat{\Pi}_N$, as follows:

$$\operatorname*{Min}_{\Pi \in \Omega_\Pi} \left\{\frac{1}{N}\sum_{t=1}^{N}(y_t - \Pi z_t)'(y_t - \Pi z_t)\right\} \equiv \frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{\Pi}_N z_t)'(y_t - \hat{\Pi}_N z_t) \,,$$

where Ω_Π is a set of matrices $-B^{-1}\Gamma$ in which B and Γ satisfy the *a priori* restrictions. We define $|\Pi|$ for any Π as

$$|\Pi| \equiv \left(\sum_{i=1}^{m}\sum_{j=1}^{n}\Pi_{ij}^2\right)^{1/2} \,,$$

where

$$(\Pi_{ij}) \equiv \Pi \,.$$

THEOREM: *Simultaneous least squares estimates $\hat{\Pi}_N$ tend to true values Π_0 in a probability sense under assumptions* (i) *through* (vi) *above; that is, for any given positive numbers ε, λ we can find a certain integer N_0 such that*

$$\Pr\{|\hat{\Pi}_N - \Pi_0| < \varepsilon\} \geqq 1 - \lambda$$

holds for any $N \geqq N_0$.

PROOF: Define $\Delta\Pi$ for any Π as

$$\Delta\Pi \equiv \Pi - \Pi_0 .$$

We have the following identity:

$$\begin{aligned} NS_N^2(Y_N, \Pi) &= \sum_{t=1}^{N} (y_t - \Pi_0 z_t - \Delta\Pi z_t)'(y_t - \Pi_0 z_t - \Delta\Pi z_t) \\ &= \sum_{t=1}^{N} (v_t - \Delta\Pi z_t)'(v_t - \Delta\Pi z_t) \\ &= \sum_{t=1}^{N} \{v_t'v_t - 2v_t'\Delta\Pi z_t + (\Delta\Pi z_t)'(\Delta\Pi z_t)\} , \end{aligned}$$

or

$$(7)\quad S_N^2(Y_N, \Pi) = S_N^2(Y_N, \Pi_0) - \frac{2}{N}\sum_{t=1}^{N} v_t'\Delta\Pi z_t + \frac{1}{N}\sum_{t=1}^{N}(\Delta\Pi z_t)'(\Delta\Pi z_t) .$$

The second term in the last expression above satisfies relations

$$\begin{aligned} \left|\frac{2}{N}\sum_{t=1}^{N} v_t'\Delta\Pi z_t\right| &= 2\left|\frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{m} v_{it}\Delta\Pi_i z_t\right| \\ &\leqq 2\sum_{i=1}^{m}\left|\Delta\Pi_i \frac{1}{N}\sum_{t=1}^{N} v_{it} z_t\right| \\ &\leqq 2\sum_{i=1}^{m}\left|\Delta\Pi_i\right|\left|\frac{1}{N}\sum_{t=1}^{N} v_{it} z_t\right| \\ &\leqq 2\,|\Delta\Pi|\sum_{i=1}^{m}\left|\frac{1}{N}\sum_{t=1}^{N} v_{it} z_t\right| \\ &= 2\,|\Delta\Pi|\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\left|\frac{1}{N}\sum_{t=1}^{N} v_{it} z_{jt}\right|^2\right\}^{1/2}, \end{aligned}$$

where v_{it} and z_{jt} denote scalar components of v_t and z_t respectively, and $\Delta\Pi_i$ denotes the i-th row vector of $\Delta\Pi$. Hence for arbitrarily given positive numbers ε and δ

$$(8)\quad \begin{aligned} &\Pr\left\{\left|\frac{2}{N}\sum_{t=1}^{N} v_t'\Delta\Pi z_t\right| < \delta, \text{ for } |\Delta\Pi| = \varepsilon\right\} \\ &\qquad \geqq \Pr\left[\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\left|\frac{1}{N}\sum_{t=1}^{N} v_{it} z_{jt}\right|^2\right\}^{1/2} < \frac{\delta}{2\varepsilon}\right]. \end{aligned}$$

On the other hand, we have

(8.1) $E(v_{it}z_{jt}|z_{j1}, z_{j2}, \cdots) = 0$,

(8.2) $\mathrm{Var}\,(v_{it}z_{jt}|z_{j1}, z_{j2}, \cdots) < \sigma^2 < \infty$,

(8.3) conditional random variables $v_{it}z_{jt}$ given $(z_{j1}, z_{j2}, \cdots)$ are independent of each other with respect to different t,

from assumptions (ii), (iii), (iv). Therefore, we get

$$\Pr\left\{\left|\frac{1}{N}\sum_{t=1}^{N} v_{it}z_{jt}\right| < \delta_0\right\} \geqq 1-\lambda_0 \qquad (i=1,\cdots,m;\ j=1,\cdots,n)$$

for any positive δ_0, λ_0, and any integer $N \geqq N_0$, taking a sufficiently large integer N_0, as a result of Bernoulli's law of large numbers. Consequently,

$$\Pr\left[\sum_{i=1}^{m}\left\{\sum_{j=1}^{n}\left|\frac{1}{N}\sum_{t=1}^{N} v_{it}z_{jt}\right|^2\right\}^{1/2} < \delta_1\right] \geqq 1-\lambda_1 \tag{9}$$

for any positive δ_1 and λ_1, and $N \geqq N_0$. Combining (8) with (9), we get the inequality

$$\Pr\left\{\left|\frac{2}{N}\sum_{t=1}^{N} v_t'\Delta\Pi z_t\right| < \delta, \text{ for } |\Delta\Pi| = \varepsilon\right\} \geqq 1-\lambda \tag{10}$$

for any positive δ and λ, and $N \geqq N_0$, taking a sufficiently large N_0.

As for last term in (7), we have following:

$$\begin{aligned}\frac{1}{N}\sum_{t=1}^{N}(\Delta\Pi z_t)'(\Delta\Pi z_t) &= \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{m}(\Delta\Pi_i z_t)^2 \\ &= \frac{1}{N}\sum_{t=1}^{N}\sum_{i=1}^{m}|z_t|^2|\Delta\Pi_i|^2\cos^2(\Delta\Pi_i, z_t) \\ &= \sum_{i=1}^{m}|\Delta\Pi_i|^2\frac{1}{N}\sum_{t=1}^{N}|z_t|^2\cos^2(\Delta\Pi_i, z_t)\,,\end{aligned}$$

where $\cos(\Delta\Pi_i, z_t)$ denotes the cosine of angle between vector $\Delta\Pi_i$ and z_t in n-dimensional Euclidian space. Since the inequality

$$\begin{aligned}&\lim_{N\to\infty}\frac{1}{N}\sum|z_t|^2\cos^2(\Delta\Pi_i, z_t) \\ &\qquad \geqq \lim_{N\to\infty}|z_t|^2\lim_{N\to\infty}\frac{1}{N}\sum\cos^2(\Delta\Pi_i, z_t) \equiv \eta_i > 0\end{aligned}$$

holds for $\Delta\Pi_i$ such that $|\Delta\Pi_i| \neq 0$, from assumptions (v) and (vi), we have

$$\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^{N}(\Delta\Pi z_t)'(\Delta\Pi z_t) \equiv f(\Delta\Pi) > 0$$

as long as $|\Delta\Pi| \neq 0$, and, considering the continuity for $f(\Delta\Pi)$ with respect to $\Delta\Pi$,

$$\inf_{|\Delta\Pi|=\varepsilon} f(\Delta\Pi) \equiv \eta > 0$$

for any positive ε. Consequently we conclude that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}(\Delta\Pi z_i)'(\Delta\Pi z_i)\geqq\eta>0 \tag{11}$$

for $|\Delta\Pi|=\varepsilon$.

We now obtain the following result from relation (7), taking δ in (10) smaller than η in (11), and combining (10) with (11):

$$\Pr\{S_N^2(Y_N,\Pi)>S_N^2(Y_N,\Pi_0),\ \text{for}\ |\Pi-\Pi_0|=\varepsilon\}\geqq 1-\lambda \tag{12}$$

for any positive λ and ε, and $N\geqq N_0$, taking a sufficiently large N_0.

If we substitute $\Pi+(k-1)\Delta\Pi$ or $\Pi_0+k\Delta\Pi$ into $S_N^2(Y_N,\Pi)$ in (7) instead of Π, where k is a scalar, we get

$$\begin{aligned}
&S_N^2(Y_N,\Pi_0+k\Delta\Pi)\\
&\quad=S_N^2(Y_N,\Pi_0)-\frac{2}{N}\sum_{i=1}^{N}v_i'k\Delta\Pi z_i+\frac{1}{N}\sum_{i=1}^{N}(k\Delta\Pi z_i)'(k\Delta\Pi z_i)\\
&\quad=S_N^2(Y_N,\Pi_0)-\frac{2}{N}k\sum_{i=1}^{N}v_i'\Delta\Pi z_i+\frac{k^2}{N}\sum_{i=1}^{N}(\Delta\Pi z_i)'(\Delta\Pi z_i)\\
&\quad=S_N^2(Y_N,\Pi_0)+k\Big\{-\frac{2}{N}\sum_{i=1}^{N}v_i'\Delta\Pi z_i+\frac{1}{N}\sum_{i=1}^{N}(\Delta\Pi z_i)'(\Delta\Pi z_i)\Big\}\\
&\qquad+\frac{k(k-1)}{N}\sum_{i=1}^{N}(\Delta\Pi z_i)'(\Delta\Pi z_i)\,.
\end{aligned}$$

Therefore, if we make $k\geqq 1$ and are concerned with a point in Y_N space such that the inequality

$$S_N^2(Y_N,\Pi)>S_N^2(Y_N,\Pi_0)$$

or the equivalent inequality

$$-\frac{2}{N}\sum_{i=1}^{N}v_i'\Delta\Pi z_i+\frac{1}{N}\sum_{i=1}^{N}(\Delta\Pi z_i)'(\Delta\Pi z_i)>0$$

holds, then inequality

$$S_N^2(Y_N,\Pi+(k-1)\Delta\Pi)>S_N^2(Y_N,\Pi_0) \tag{13}$$

must hold for the same point of Y_N space and the translated point of Π, which can be written $\Pi+(k-1)(\Pi-\Pi_0)$. However, any point in Π space for which $|\Pi_0-\Pi|\geqq\varepsilon$ can be expressed as $\Pi_s+(k-1)(\Pi_s-\Pi_0)$, using a point Π_s such that $|\Pi_s-\Pi_0|=\varepsilon$, and a scalar k such that $k\geqq 1$. So if the inequality

$$S_N^2(Y_N,\Pi)>S_N^2(Y_N,\Pi_0)$$

holds for every point in a certain set of Y_N space, and for every point of Π space such that $|\Pi-\Pi_0|=\varepsilon$, then the same inequality

must hold for every point in the same set of Y_N space, and for every point of Π space such that $|\Pi - \Pi_0| \geqq \varepsilon$. As a result we get

$$\begin{aligned}(14)\qquad &\Pr\{S_N^2(Y_N, \Pi) > S_N^2(Y_N, \Pi_0), \text{ for } |\Pi - \Pi_0| = \varepsilon\} \\ &= \Pr\{S_N^2(Y_N, \Pi) > S_N^2(Y_N, \Pi_0), \text{ for } |\Pi - \Pi_0| \geqq \varepsilon\}\,.\end{aligned}$$

Thus we derive the conclusion from (12) and (14) that

$$(15)\qquad \Pr\{S_N^2(Y_N, \Pi) > S_N^2(Y_N, \Pi_0), \text{ for } |\Pi - \Pi_0| \geqq \varepsilon\} \geqq (1 - \lambda)\,,$$

which holds for any given positive λ, ε, and $N \geqq N_0$, taking a sufficiently large integer N_0.

In another context the inequality (15) can be transformed into

$$\Pr\{|\hat{\Pi}_N - \Pi_0| < \varepsilon\} \geqq 1 - \lambda\,,$$

because the fact that the inequality $S_N^2(Y_N, \Pi) > S_N^2(Y_N, \Pi_0)$ holds for all Π satisfying $|\Pi - \Pi_0| \geqq \varepsilon$, for a particular point of Y_N space, implies that every point of Π space satisfying $|\Pi - \Pi_0| \geqq \varepsilon$ cannot qualify as $\hat{\Pi}_N$, given the point of Y_N space which occurs.

From assumption (i), Π_0 and the consistent estimate $\hat{\Pi}_N$ of Π_0, obtained under the restrictions $\hat{\Pi}_N \in \Omega_\Pi$, can be uniquely transformed into B_0, Γ_0 and the estimates of B_0, Γ_0 respectively, through *a priori* information about B and Γ. Accordingly, estimates thus transformed necessarily become consistent estimates of the original structural parameters, since this transformation must be continuous with respect to Π, because of its uniqueness and the continuous mapping of B and Γ into Π.

Finally, it may be remarked that if one is interested in the theorem only, the proof can be substantially simplified by starting from the consistency property of simple least squares estimation of the reduced forms.

University of Pennsylvania, U.S.A., and
Tokyo Metropolitan University, Japan

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3, September, 1960

PROBLEMS IN PRODUCTION-INVESTMENT PLANNING OVER TIME*

BY VERNON L. SMITH[1]

1. INTRODUCTION

IN A PREVIOUS PAPER [5] I attempted to develop the foundations for an integrated theory of production and investment in which the choice of best input-output technique, including the optimal stock levels of durable inputs, and the optimal "replacement" of durable inputs were simultaneously determined in a cost minimization model. Very little was done in that paper to develop the dynamics of the theory, which is the objective of the present paper.

We imagine individual decision making units, which we call "firms," whose input-output decisions are constrained by a convex function relating the stock and flow inputs to a producing process. This function involves both flow and physical stock inputs because the physical stock of the typical durable input to production directly influences the transformation function of flow inputs into flow outputs.[2] A most important characteristic of such individual durable goods is that they can be assumed to be continuously variable in size when considering expansionary investment decisions but cannot be contracted in amount, except by shutting down individual physical units. Typical examples are boilers, buildings, pumps, and pipe lines. For example, a steam generating plant can install any size boiler, X, measured, say, in square feet of heating surface. But once a unit of size X^0 is installed, disinvestment can occur only by removing the unit entirely from production. Once a unit has been installed, expansion can take place either by adding a parallel unit of appropriate size or discarding the existing facility and replacing it with a larger unit. Hence, the decision problem contains a curious mixture of continuity and lumpiness.

The analysis will be confined to processes which require one current input and one capital equipment input.

* Manuscript received January 15, 1960; revised April 26, 1960.

[1] The research underlying this paper was supported by a Ford Foundation Faculty Research Fellowship, 1958-1959. I am indebted to the referee for his numerous detailed suggestions and corrections which have materially improved the mathematical exposition and analysis of the paper.

[2] *Ibid.*, [5, pp. 62-66].

2. THE SINGLE INVESTMENT

Suppose the firm's planning horizon is T years. Assume that a single investment decision is to be made over the planning horizon, and that decision concerns the initial stock of capital to be installed by the firm. Assume further that output requirements as a continuous function of time are known with certainty, say $y(t), 0 \leq t \leq T$. The firm must choose the level of investment outlay to be made initially and the rate of consumption of current input at each instant over the horizon. The consumption of the current input at time t is $x_1(t)$, while the quantity of the capital good to be installed is X_2. The process is constrained by a production function of the form $y(t) = f[x_1(t), X_2]$ which exhibits non-decreasing returns to scale. It is assumed that no output can be produced without some positive amount of each input. If the firm desires to minimize discounted total current account cost[3] over the interval $0 \leq t \leq T$, then the problem is to choose $x_1(t) \geq 0$ and $X_2 \geq 0$ so as to minimize

$$C = \frac{r}{1 - e^{-rT}} \int_0^T w_1 x_1(t) e^{-rt} dt + \frac{r W_2 X_2}{1 - e^{-rT}} , \tag{2.1}$$

subject to

$$y(t) = f[x_1(t), X_2] , \tag{2.2}$$

where w_1 and W_2 are the prices of the current and capital inputs, and r is the rate of interest.

An equivalent problem is to minimize

$$\varphi = \frac{r}{1 - e^{-rT}} \int_0^T (w_1 x_1(t) e^{-rt} - \lambda \{ f[x_1(t), X_2] - y(t) \}) dt + \frac{r W_2 X_2}{1 - e^{-rT}} \tag{2.3}$$

with respect to the *extremal* $x_1(t)$ and the *variable* X_2, where λ is a function of w_1, W_2, T, r, $y_0^T(t)$, and t, to be determined shortly. The function λ plays a role entirely analogous to the Lagrange multiplier of the ordinary theory of relative optima.

Note that the expression for φ in (2.3) is a special form of the following standard calculus of variations problem: Minimize

[3] "Discounted total current account cost," or simply "current account cost," as I call it, is the constant outlay stream that has a present value equal to that of all actual outlays for current and capital inputs over the horizon. For the models in this paper, minimizing current account cost gives the same decision rules as minimizing the present value of all outlays.

$$(2.3') \qquad I = \int_0^t F(x, \dot{x}, t)dt \,,$$

where $x = x(t)$ and $\dot{x} = dx/dt$.[4] The solution to (2.3') is contained in the Euler equation $\frac{\partial F}{\partial x} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}}\right) = 0$ in $x(t)$. We apply this condition directly to (2.3) in minimizing φ with respect to the extremal $x_1(t)$, and we apply the condition $\partial\varphi/\partial X_2 = 0$ in minimizing with respect to the variable X_2. If we let F stand for the integrand in (2.3), necessary conditions for an interior minimum of φ are[5]

$$(2.4) \qquad \frac{\partial F}{\partial x_1} = w_1 e^{-rt} - \lambda f_1[x_1(t), X_2] = 0 \,,$$

$$(2.5) \qquad \frac{\partial \varphi}{\partial X_2} = -\frac{r}{1 - e^{-rT}}\int_0^T \lambda f_2[x_1(t), X_2]dt + \frac{rW_2}{1 - e^{-rT}} = 0 \,,$$

where $f_1 = \partial f/\partial x_1$ and $f_2 = \partial f/\partial X_2$, and a similar conventional notation is used throughout this paper.

From the production function (2.2) and the necessary conditions (2.4) and (2.5) we determine (if they exist) the optimal consumption path of current input $x_1^0(t)$, the optimal level of investment, X_2^0, and, as a by-product, the function $\lambda^0 = \lambda(w_1, W_2, T, r, y_0^T(t), t)$ is determined. Specifically, by eliminating λ from (2.4) and (2.5), we get

$$(2.6) \qquad -\frac{r}{1 - e^{-rT}}\int_0^T \frac{w_1 f_2[x_1(t), X_2]}{f_1[x_1(t), X_2]} e^{-rt}dt + \frac{rW_2}{1 - e^{-rT}} = 0 \,.$$

The interpretation of this expression is easily clarified. The ratio f_2/f_1 is the amount by which the consumption of input number 1 is decreased (increased) at time t if the initial installation of capital is increased (decreased) infinitesimally. Then $w_1 f_2/f_1$ is the rate of sav-

[4] See, e.g., Fox [2], pp. 3-34. Also see Fox [2], chapter IV, *passim*, especially pp. 80-83 and pp. 94-96, for a discussion of constrained extremal problems where the subsidiary equation is of nonintegral type. The problem presented by Fox (in one variable form) is to maximize or minimize $I = \int_{t_1}^{t_2} G(x, \dot{x}, t)dt$ subject to $S(x, \dot{x}, t) = 0$. He then introduces λ, a function of t to be determined, and develops a solution (*ibid.*, [2], p. 96) of the form $\frac{\partial G}{\partial x} - \frac{d}{dt}\left(\frac{\partial G}{\partial \dot{x}}\right) - \left\{\lambda\frac{\partial S}{\partial x} - \frac{d}{dt}\left(\lambda\frac{\partial S}{\partial \dot{x}}\right)\right\} = 0$. This is equivalent to the solution $\frac{\partial F}{\partial x} - \left(\frac{d}{dt}\right)\frac{\partial F}{\partial \dot{x}} = 0$ with $F = G(x, \dot{x}, t) - \lambda S(x, \dot{x}, t)$. The well-known Lagrange multiplier shadow price tricks of static theory apply to our class of calculus of variations problems.

[5] Note that the condition (2.4) can be obtained by straight scalar optimization, since the integral in (2.3) does not contain any time derivatives of the extremal. The calculus of variations form of the solution is emphasized because it underlies all the problems to be discussed, and some of the subsequent problems cannot be treated by scalar techniques.

ings in the cost of the current input per unit increase in the initial stock of capital. Hence, the integral of $(w_1 f_2/f_1)e^{-rt}$ over the interval $0 \leq t \leq T$ is the present value of the savings in current input cost per unit increase in capital investment. Equation (2.6) states that investment is expanded, given the time pattern of output requirements, until the discounted annual value of the current input cost saved from the last increment of capital equals the unit rent of that increment of capital, $rW_2/(1 - e^{-rT})$. Now, by integrating (2.6) in combination with (2.2), we determine the optimal values $x_1^0(t)$ and X_2^0. With $x_1^0(t)$ and X_2^0 determined, (2.4) provides the Lagrangean function λ^0. Hence, an interior solution to (2.2), (2.4), and (2.5) can be written

$$x_1^0(t) = h_1(w_1, W_2, T, r, y_0^T(t), t) , \tag{2.7}$$

$$X_2^0 = H_2(w_1, W_2, T, r, y_0^T(t)) , \tag{2.8}$$

$$\lambda^0 = \lambda(w_1, W_2, T, r, y_0^T(t), t) . \tag{2.9}$$

The similarity between this solution and the analogous static solution will be evident.

Can we carry the analogy with static theory a step further and interpret λ as a marginal cost function? The answer is "yes". As is evident from equation (2.4), when put in the form $\lambda = w_1 e^{-rt}/f_1[x_1(t), X_2]$, λ *is the discounted instantaneous* (or short-run) *marginal operating cost function*. The effect of changes in w_1 or W_2 on the consumption of the current input $x_1^0(t)$ and the level of investment X_2^0 parallel the effect of similar changes in the analogous static model. The higher the price of capital relative to the price of current input, the lower is investment and the higher the consumption of current input.

3. THE TWO INVESTMENTS CHAIN

Consider a slight extension of the problem discussed in the last section. Assume again a finite horizon of T years. with a known dynamic requirements function, $y(t)$, defined over the interval $0 \leq t \leq T$. We divide the interval into two parts of unknown lengths, T_1 and T_2, thus forming two operating periods, I and II, as shown in Figure 1. It is assumed that *at most* two investments are to be made in the interval $0 \leq t \leq T$. The first investment is at $t = 0$, the second at $t = T_1$, to be determined. We wish to determine conditions for the following: (1) how much capital to install initially; (2) how much capital to install at T_1; (3) when to install (if at all) the second capital facility, i.e., the determination of T_1; (4) the optimal instantaneous operating rules in periods I and II; (5) the optimal consumption path for the

current input in periods I and II; and (6) rules for discarding either capital facility in period II if such action is optimal. To continue with the earlier example, we want to know how large a boiler to install now, and how large a boiler to install at any later time. We also want to know how to load the boiler or boilers at each instant over the interval, how much current input (fuel) will be consumed at each instant, and when, if at all, to discard either boiler.

We assume momentarily that neither facility is to be idled temporarily over any subinterval. If at any time it pays to withdraw one facility from production, it is assumed that the action is permanent and the facility is discarded at zero salvage value.

Using a current account cost criterion, the objective is to minimize

$$C = \frac{r}{1 - e^{-rT}}\left\{\int_0^{T_1} w_1 x_{11}^1(t) e^{-rt} dt + W_2 X_{21}\right\} + \frac{r}{1 - e^{-rT}}\left\{\int_{T_1}^{T} w_1 [x_{11}^2(t) + x_{12}^2(t)] e^{-rt} dt + W_2 X_{22} e^{-rT_1}\right\}, \tag{3.1}$$

where X_{21} is the quantity of capital input installed at the beginning of operating period I (facility number 1), X_{22} is the amount installed at the beginning of period II (facility number 2), $x_{11}^1(t)$ is the consumption path of current input by facility 1 in period I, $x_{11}^2(t)$ is the consumption path of current input by facility 1 in period II, and $x_{12}^2(t)$ is

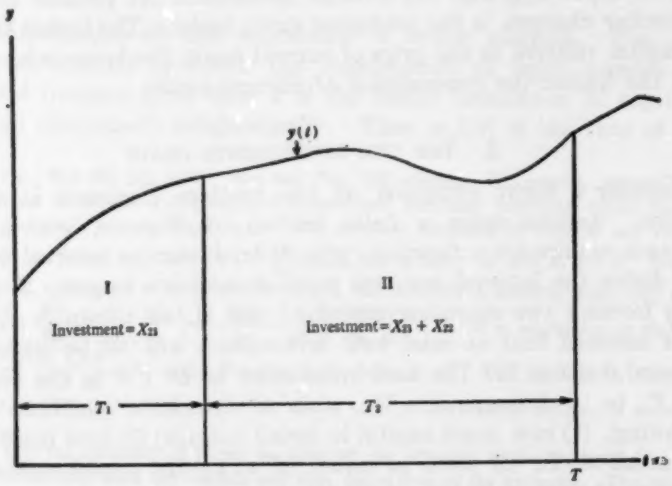

FIGURE 1

the consumption path of current input by facility 2 in period II. The expression (3.1) is to be minimized subject to

$$y(t) = f^1[x^1_{11}(t), X_{21}], \qquad 0 \leq t < T_1, \tag{3.2}$$

and

$$y(t) = f^1[x^1_{11}(t), X_{21}] + f^2[x^2_{12}(t), X_{22}], \qquad T_1 \leq t \leq T. \tag{3.3}$$

The functions f^1 and f^2 refer to the production function for facilities 1 and 2. That is, the production technology ruling at the beginning of period I is $f^1[x^1_{11}(t), X_{21}]$, while the production technology available at the beginning of period II is $f^2[x^2_{12}(t), X_{22}]$. By time T_1, technological improvements in equipment and operating technique may have rendered f^2 entirely dissimilar to f^1. We allow for this by permitting the production function to alter from one investment point to the next.[6]

By introducing two Lagrangean functions λ_1 and λ_2, one for each of the constraints in (3.2) and (3.3), a problem equivalent to the one just stated is to minimize

$$\begin{aligned}\varphi = {} & \frac{r}{1-e^{-rT}}\left\{\int_0^{T_1}\left(w_1 x^1_{11}(t)e^{-rt} - \lambda_1[f^1[x^1_{11}(t), X_{21}] - y(t)]\right)dt + W_2 X_{21}\right\} \\ & + \frac{r}{1-e^{-rT}}\left\{\int_{T_1}^{T}\left(w_1[x^2_{11}(t) + x^2_{12}(t)]e^{-rt} - \lambda_2[f^1[x^2_{11}(t), X_{21}]\right.\right. \\ & \left.\left. + f^2[x^2_{12}(t), X_{22}] - y(t)]\right)dt + W_2 X_{22}e^{-rT_1}\right\},\end{aligned} \tag{3.4}$$

with respect to the nonnegative extremals $x^1_{11}(t)$, $x^2_{11}(t)$, $x^2_{12}(t)$, and the nonnegative variables X_{21}, X_{22}, T_1 (since T is fixed, T_2 is determined once we know T_1).

Assuming an interior solution, for the present, the necessary *intratemporal* conditions for minimum φ with respect to the current input consumption paths $x^1_{11}(t)$, $x^2_{11}(t)$, and $x^2_{12}(t)$, in that order, are

$$w_1 e^{-rt} - \lambda_1 f^1_1[x^1_{11}(t), X_{21}] = 0, \qquad 0 \leq t < T_1 \tag{3.5}$$

$$\left.\begin{aligned} w_1 e^{-rt} - \lambda_2 f^1_1[x^2_{11}(t), X_{21}] &= 0, \\ w_1 e^{-rt} - \lambda_2 f^2_1[x^2_{12}(t), X_{22}] &= 0. \end{aligned}\right\} \qquad T_1 \leq t \leq T_1 + T_2 \tag{3.6, 3.7}$$

The *intertemporal* conditions on the investment levels X_{21} and X_{22} are

[6] A more general way to introduce technological change is to write the second period production function in the form $f^2[x^2_{12}(t), X_{22}, T_1]$, i.e., the "state of the arts" is a function of the point in calendar time at which the capital equipment of a process is installed. The class of problems under discussion is tedious enough without pursuing maximum generality

$$-\frac{r}{1-e^{-rT}}\left\{\int_0^{T_1}\lambda_1 f_2^1[x_{11}^1(t), X_{21}]dt + \int_{T_1}^{T}\lambda_2 f_2^1[x_{11}^2(t), X_{21}]dt - W_2\right\} = 0\,, \tag{3.8}$$

$$-\frac{r}{1-e^{-rT}}\left\{\int_{T_1}^{T}\lambda_2 f_2^2[x_{12}^2(t), X_{22}]dt - W_2 e^{-rT_1}\right\} = 0\,. \tag{3.9}$$

Finally, the optimal timing of the second investment, T_1, must satisfy the condition $\partial\varphi/\partial T_1 = 0$, or

$$w_1 x_{11}^1(T_1) - w_1[x_{11}^2(T_1) + x_{12}^2(T_1)] - rW_2X_{22} = 0\,. \tag{3.10}$$

The conditions (3.5)–(3.10) contain practically the whole of a modern, marginalist, dynamic theory of production and investment. λ_1 in (3.5) and (3.8) is the discounted marginal operating cost function in period I, i.e., $\lambda_1 = w_1e^{-rt}/f_1^1$ over the interval $0 \leq t < T_1$. λ_2 is the discounted marginal operating cost function in period II. From (3.6) and (3.7) we have $\lambda_2 = w_1e^{-rt}/f_1^1 = w_1e^{-rt}/f_1^2$, for $T_1 \leq t \leq T$, during operating period II, i.e., the two facilities are operated at output (and current input) rates that equalize their instantaneous marginal costs (and marginal productivities).

In interpreting equation (3.8), note that the first integral under the brackets gives the present value of the period I savings in current input resulting from an infinitesimal increase in the initial investment level X_{21}. Similarly, the second integral gives the value of the corresponding period II savings. Therefore (3.8) requires the initial investment to be expanded until the discounted annual value of the savings from the last unit of capital equals the unit rent of capital, $rW_2/(1-e^{-rT})$. Equation (3.9) requires the level of additional investment at T_1 to be expanded until the annual discounted value of the savings resulting from the last unit of that capital is equal to the unit rent of capital.

The first term in (3.10) is the instantaneous operating cost, at T_1, of operating facility number 1 alone. The second term is the initial cost of operating facilities number 1 and 2 in parallel after the latter has been installed. Therefore, the difference between these terms is the initial net rate of operating cost savings resulting from adding the new facility. The last term in (3.10) is the instantaneous interest on the investment in a new facility. Therefore (3.10) says to time the installation of a second parallel facility when the rate of operating cost savings equals the instantaneous interest on the new investment. In a word, a new facility should be installed whenever the first "year" savings effected by the new unit will pay the first "year" of interest on the new investment.

So far the analysis has assumed an interior solution. Can we replace the strict equalities in (3.5)–(3.10) with inequality signs and give Kuhn-Tucker interpretations to the surplus fulfillment of one or more of these conditions? We can (see, e.g., Morin [4], p. 448), and the resulting conditions will provide rules for an optimal equipment discard policy.

Consider first the conditions (3.5)–(3.7), which we rewrite in the form

$$w_1 e^{-rt} - \lambda_1 f_1^1[x_{11}^1(t), X_{21}] \geq 0, \qquad 0 \leq t < T_1 \tag{3.5'}$$

$$\left.\begin{aligned} w_1 e^{-rt} - \lambda_2 f_1^1[x_{11}^2(t), X_{21}] &\geq 0, \\ w_1 e^{-rt} - \lambda_2 f_1^2[x_{12}^2(t), X_{22}] &\geq 0, \end{aligned}\right\} \quad T_1 \leq t \leq T. \tag{3.6'}, \tag{3.7'}$$

If ">" holds in (3.5') for some $t = t_{11}$, $0 \leq t_{11} < T_1$, then $x_{11}^1(t_{11}) = 0$, i.e., at t_{11} it pays to withdraw the first facility from production. If this occurs in the first operating period it means that output requirements have declined to zero. If ">" holds in (3.6') for some $t = t_{12}$, $T_1 \leq t_{12} \leq T$, then $x_{11}^2(t_{12}) = 0$, and at t_{12} the first facility is withdrawn from production in favor of the singular operation of the new facility installed at T_1. Or, if ">" holds in (3.7') for some t_{22}, $T_1 \leq t_{22} \leq T$, then $x_{12}^2(t_{22}) = 0$, and at t_{22} the second facility is withdrawn from production leaving the original facility only. Hence, by proper interpretation of the corner solutions implied by (3.5')–(3.7'), we can obtain all possible equipment discard alternatives contained in the planning interval $0 \leq t \leq T$. Note the strategic role played by the inequalities; they define implicitly the firm's optimal equipment discard policies. Our formulation of the problem allowed explicitly *for equipment purchases only*, i.e., at 0 and T_1. The discard points, if they exist, are then evolved by interpretation of the equilibrium condition. Also note that the extension of our analysis to dynamic problems has not destroyed the intuitive usefulness of the Lagrange multipliers; they continue to provide valuable service as shadow price evaluators. In this sense the conditions (3.5')–(3.7') require each facility to be operated at all times so that the internal instantaneous value of a unit of the current input, e.g., $\lambda_1 f_1^1[x_{11}^1(t), X_{21}]e^{rt}$ in (3.5'), does not exceed the external cost, w_1, of acquiring a unit of this input.

Similarly, we can replace "=" with "$\geq$" in (3.8) and (3.9) to obtain more general conditions on the investment levels X_{21} and X_{22}. If ">" holds in (3.8), then $X_{21} = 0$. This could only obtain if requirements are zero throughout the first operating period, since we assume that no output can be produced without some positive investment. If ">" holds in (3.9), then $X_{22} = 0$, and the best policy is to employ singly the original facility, $X_{21} > 0$, throughout the planning period.

The condition (3.10) could also be written with "$\geq$" in place of "$=$". If "$>$" holds, then $T_1 = 0$, and only a single facility is used over the entire period. Hence, $X_{21} = 0$ and $T_1 = 0$ have equivalent implications.

What happens if we drop the assumption that neither facility is to be idled temporarily over any subinterval? We can still write conditions on optimal policies, but the problem becomes more cumbersome. Now, if "$>$" holds for some t in any of the conditions (3.5′)–(3.7′), the corresponding facility is discarded only if there is no later $t' > t$ at which the "$=$" in that condition can be satisfied. For example, suppose "$>$" holds in (3.6′) for some interval $T_1 \leq t_{12} \leq t \leq t'_{12} \leq T$, with the the equality holding elsewhere. Then at t_{12}, instead of discarding the equipment, it is placed in standby until t'_{12}, at which time it is returned to the operating state.

4. THE MULTIPLE INVESTMENT CHAIN

Turning next to a generalization of the two-input problem in production-investment dynamics, our objective is to extend the analysis of the last section to the infinite horizon investment chain. We divide the planning interval, $0 \leq t \leq \infty$, into an infinite sequence of segments of unknown lengths $T_1, T_2, T_3, \cdots$, forming an infinite sequence of operating intervals, as shown in Figure 2. The objective in this case is to minimize the discounted current account cost over all future operating intervals, subject to input-output constraints of the form (3.2) and (3.3) for each operating interval. By introducing Lagrange

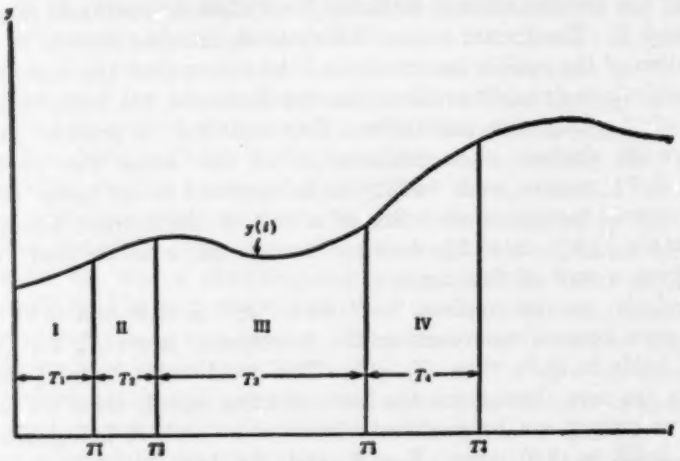

FIGURE 2

multipliers $\lambda_1, \lambda_2, \lambda_3, \cdots$, this constrained minimum problem is equivalent to minimizing the tedious expression

$$\varphi = r\sum_{k=1}^{\infty}\left\{\int_{T^*_{k-1}}^{T^*_k}\left(w_1\sum_{j=1}^{k}x^k_{1j}(t)e^{-rt} - \lambda_k\left[\sum_{j=1}^{k}f^j[x^k_{1j}(t), X_{2j}] - y(t)\right]\right)dt + W_2X_{2k}e^{-rT^*_{k-1}}\right\}, \tag{4.1}$$

where $T^*_k = \sum_{i=1}^{k} T_i$, and $T^*_0 = 0$, with respect to the extremals $x^k_{1j}(t)$, the capital stock levels X_{2k}, and the purchasing dates $T^*_k (j \leq k = 1, 2, 3, \cdots)$. The expression (4.1) is the infinite dimensional counterpart of (3.4). We again assume that whenever optimality requires a facility to be idled, that facility is discarded.

With respect to the extremals, necessary conditions for minimizing (4.1) are

$$w_1e^{-rt} - \lambda_k f^i_1[x^k_{1i}(t), X_{2i}] \geq 0, \qquad k = 1, 2, 3, \cdots; \quad i = 1, 2, \cdots, k. \tag{4.2}$$

The infinite dimensional conditions on the X_{2k} and T_k are

$$-r\sum_{k=j}^{\infty}\left\{\int_{T^*_{k-1}}^{T^*_k}\lambda_k f^j_1[x^k_{1j}, X_{2j}]dt\right\} + re^{-rT^*_{j-1}}W_2 \geq 0, \qquad j = 1, 2, 3, \cdots \tag{4.3}$$

and

$$\sum_{k=j}^{\infty}w_1\sum_{h=1}^{k}x^k_{1h}(T^*_k)e^{-rT^*_k} - \sum_{k=j+1}^{\infty}w_1\sum_{h=1}^{k}x^k_{1h}(T^*_{k-1})e^{-rT^*_{k-1}} - r\sum_{k=j+1}^{\infty}W_2X_{2k}e^{-rT^*_{k-1}} \geq 0, \qquad j = 1, 2, 3, \cdots \tag{4.4}$$

where $T^*_k = \sum_{i=1}^{k} T_i$ and $T^*_0 = 0$. The intratemporal conditions (4.2) require the current input to each facility in all operating periods to have a marginal inputed value which does not exceed the market price of the input. When the equalities hold, these conditions also imply that all facilities in any period must be operated at output levels which equate their marginal operating costs (and marginal physical productivities). If, at any instant in any period, an equality contained in (4.2) cannot be maintained, the appropriate facility is at that moment idled and, therefore, by assumption, discarded.[7]

The intertemporal conditions (4.3) require investment in a new facility at each purchase date to be expanded until the discounted current account cost of all future savings in operating expenses, effected by

[7] The problem of temporary idling can be handled by seeking subintervals over which the appropriate inequalities hold in the conditions (4.2).

the last increment of physical investment, is equal to the annual rent of the capital good. The integral in (4.3) serves the function of cumulating these discounted savings in each future operating period. These savings are then summed to obtain their total. If the rent of capital exceeds these savings for all feasible investment levels (the ">" prevails), then the investment is not undertaken. It is entirely possible that these conditions will signal an end to the investment chain. For example, if optimality requires the nth facility to be the last in the chain, then ">" will hold in (4.3) for X_{2k}, $k = n+1$, $n+2, n+3, \cdots$. The same situation will also be reflected in (4.4) if there is no solution for T_{n+1}, i.e., no new facilities are acquired after the nth investment. In general, the conditions (4.4) determine the length of the operating intervals $T_1, T_2, T_3, \cdots$, and therefore the purchase dates $T_1^*, T_2^*, T_3^*, \cdots$. These conditions require a new parallel facility to be installed whenever the discounted cost of operating the existing complex of parallel facilities is equal to or exceeds the discounted annual operating and capital cost of an additional facility. Each year we imagine the firm comparing the cost of maintaining the status quo (i.e., continuing to operate the existing equipment complex without benefit of an additional unit) with the cost of adding a new unit, proper account being taken of the savings effected by the latter. The status quo is maintained an additional year if the discounted cost of making the addition exceeds that of not making it. Abrupt technological improvements, reflected by a shift in new equipment production functions, are, of course, favorable to the early purchase of a new unit.

5. EXTENSIONS AND MORE GENERAL MODELS

a. *Discussion.* The solutions of the last section are quite general for the two-input case. To obtain more specific results requires more specific information concerning output requirements, price parameters, and technological change. For example, in Figure 3, assuming an infinite horizon, the solution might require a chain of five investment expansions to be made at A, B, C, D, and E. The third new facility might be added at C, even though output requirements have risen only slightly since the second facility was installed at B, because of technological improvements occurring after the second facility was purchased. The investment chain might end at E, after five facilities have been purchased, if output requirements are due to decline to a constant level as shown, and no important technological improvements are in store. At F, t_{15} years after the purchase of the last facility, the declining

output level may precipitate the discard of the first facility (or any of the other five facilities depending upon the state of technology at the time each was purchased and the scale of the unit purchased).

Can we generalize our two-input results to multiple inputs? The extension is easy if we have multiple *current* inputs, with a single durable input. The solution is identical with that of (4.2)–(4.4), except that we must add conditions of the form (4.2) for each current input. Such additional conditions simply require every current input to be consumed at rates which equalize their marginal costs, i.e., the ratio of the price of each input to the marginal productivity of that input must equal the ratio for every other current input in each operating period. The case of multiple durable inputs is not so simple. It is easy to see intuitively that conditions like those of (4.3) must surely apply to every type of durable input to the process. That is, at every purchase date for each type of equipment, investment is expanded until the savings on both current and capital account of the last increment of the capital good are equal to its annual rent. But to deduce such conditions in a form similar to (4.3) is most formidable, since the moment we admit more than a single capital good we lose the one-to-one correspondence between the equipment purchase dates $T_1^*, T_2^*, T_3^*, \cdots$, and the operating intervals $T_1, T_2, T_3, \cdots$. That is, we must now distinguish purchase dates for each equipment type. In the general case of n current inputs and m capital inputs, we cannot even write an expression of the form (4.1) without knowing how the

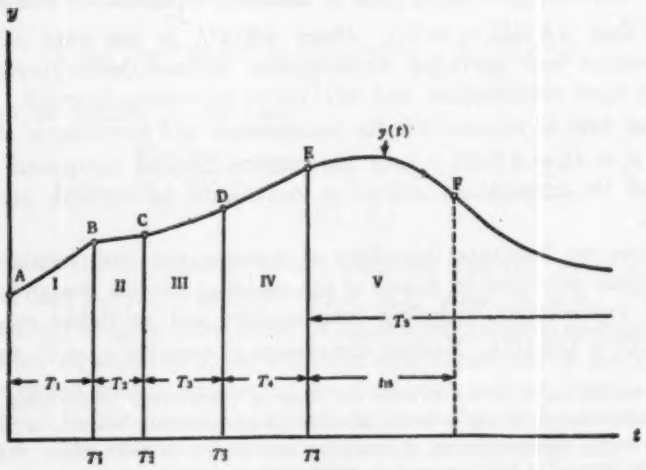

FIGURE 3

purchase dates for each capital input are ordered with respect to each other. Furthermore, the problem cannot be side-stepped by attempting to by-pass the production function and formulating the investment problem directly in terms of a current account cost function. There still remains the problem of determining such a cost function, and defining the optimal policies which it contains implicitly. The hypothesis of a cost function always presumes a wise class of optimal decisions to have been made. Also, the problem is not solved by formulating it in terms of "capacity" in the sense of output. This is a most nebulous concept, and leaves unanswered the question of the connection between "capacity" and the stocks of equipment in kind which are the direct objects of adjustment in investment decisions.

b. *The Solution with Maintenance Outlays.* There are, no doubt, many different directions in which the dynamic models which we have developed can be extended. One important element which has not been incorporated into the dynamic analysis is the cost of capital equipment maintenance and servicing. It will be assumed that the rate of increase in maintenance costs (per unit increase in the consumption of the current input) is an increasing function of the cumulative consumption of current input. I have argued elsewhere [5], pp. 78–80 that equipment maintenance cost can be expected to vary with the equipment's cumulative consumption of fuel or energy.[8] The latter is a direct measure of the total exposure of a machine to wear and tear. If we let $U_1 = U_1(t)$ be the cumulative consumption of the current input up to time t, then $U_1 = \int_0^t x_1(\tau)d\tau$. Our assumption is that $dM/dU_1 = m(U_1)$, where dM/dU_1 is the rate at which maintenance and servicing expenditures increase with increases in current input consumption, and $m(U_1)$ is an increasing function. Hence, the time rate of expenditure for maintenance and servicing is $\dot{U}_1 m(U_1)$, where $x_1 = \dot{U}_1 = dU_1/dt$. Both the current rate of equipment utilization and its cumulative utilization contribute to current maintance outlays.

Suppose we illustrate the effect of maintenance cost considerations on dynamic solutions by means of the model of Section 2 above. Total current expenditure, including maintenance and servicing outlays, is now $w_1\dot{U}_1 + \dot{U}_1 m(U_1)$. Making the change of variable $x_1 = \dot{U}_1$ through-

[8] Also see [6] pp. 26-28 for an empirical study of truck-tractor maintenance costs, in which maintenance costs per mile are found to be an increasing function of cumulative mileage. In the present context, it would be argued that a more direct explanatory variable is cumulative fuel consumption rather than cumulative mileage.

out, the expression corresponding to (2.3) to be minimized can be written

$$\varphi = \frac{r}{1-e^{-rT}}\int_0^T([w_1\dot{U}_1 + \dot{U}_1 m(U_1)]e^{-rt} - \lambda\{f(\dot{U}_1, X_2) - y(t)\})dt + \frac{rW_2X_2}{1-e^{-rT}}. \tag{5.1}$$

The integral in (5.1) is of the form (2.3) and is minimized with respect to the extremal $U_1(t)$ if $\frac{\partial F}{\partial U_1} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{U}_1}\right) = 0$. In this case the derivative with respect to $\dot{U}_1$ will enter the solution, whereas in our previous analysis the derivative with respect to the time rate of change of the extremal was identically zero. In this model we get a true calculus of variation problem, whereas previously the analysis reduced to ordinary scalar calculus minimization.

Equation (5.1) is therefore minimized with respect to $U_1(t)$ if the following differential equation is satisfied:

$$\dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt} - \frac{d}{dt}\{[w_1 + m(U_1)]e^{-rt} - \lambda f_1(\dot{U}_1, X_2)\} = 0. \tag{5.2}$$

Integrating (5.2) and solving for λ gives

$$\lambda(t) = \frac{[w_1 + m(U_1)]e^{-rt} - \int_0^t \dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt}dt + C_1}{f_1(\dot{U}_1, X_2)}, \tag{5.3}$$

where C_1 is an arbitrary constant of integration. Since

$$\int_0^T \dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt}dt = \int_0^t \dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt}dt + \int_t^T \dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt}dt,$$

(5.3) can be written in the form

$$\lambda(t) = \frac{\int_0^T \dot{U}_1\frac{\partial m}{\partial U_1}e^{-rt}dt + [w_1 + m(U_1)]e^{-rt} + K_1}{f_1(\dot{U}_1, X_2)}. \tag{5.4}$$

If at $t = T$, the termination of the planning interval, we impose the condition $\lambda(T) = [w_1 + m(U_1)]e^{-rT}/f_1(\dot{U}_1, X_2)$, then $K_1 = 0$. The Lagrangean $\lambda(t)$ is again interpreted as discounted marginal operating cost at time t. Equation (5.4) says that in equilibrium $\lambda(t)$ is the ratio of the discounted cost of employing an additional unit of input number 1 at time t to the marginal physical productivity of input number 1 at time t. The discounted cost of employing an additional unit of input number 1 at time t is composed of three parts. The first part,

$\int_t^T \dot{U}_1(\partial m/\partial U_1)e^{-rt}dt$, is the additional cost of maintaining the stock of capital over the rest of the planning interval from t to T as a result of consuming an additional unit of input number 1 at t. The second part, w_1e^{-rt}, is the direct cost of buying the additional input. The third part, $m(U_1)e^{-rt}$, is the additional cost of maintaining the capital in the current period, i.e. at t, when an additional unit of current input is consumed at t. (Remember that the consumption of input number 1 exerts both a current and a cumulative influence on maintenance outlays.)

Finally, by setting $\partial\varphi/\partial X_1 = 0$, we get the condition

$$\frac{r}{1-e^{-rT}}\int_0^T -\lambda(t)f_1(\dot{U}_1, X_1)dt + \frac{rW_1}{1-e^{-rT}} = 0\,. \tag{5.5}$$

Equation (5.4) and (5.5), together with the constraint $y(t) = f(\dot{U}_1, X_1)$, provide necessary conditions for determining x_1^0, X_1^0, and λ^0 over the interval $0 \leq t \leq T$.

We have written the solution with maintenance outlays for the case of a single investment and finite horizon. The solution is easily extended to a chain of investments over a finite or infinite horizon. Where the Kuhn-Tucker-Euler conditions require a facility to be idled (or discarded), we get $x_1(t) = \dot{U}_1(t) = 0$, and total operating cost $w_1\dot{U}_1 + \dot{U}_1m(U_1)$ drops to zero for the idled facility.

c. *Inventory Policy and Production Planning; An Extension of the Modigliani-Hohn Model.* An alternative extension of our fundamental models is to drop the implicit assumption of a nonstorable product. This introduces the necessity for simultaneous treatment of optimal inventory policy and optimal production-investment technique. We will follow a continuous version of the Modigliani-Hohn [3] model[9] in introducing inventory considerations into our analysis. Modigliani and Hohn assume a given marginal production cost function which depends only on the rate of production at time t. In our analysis the short-run marginal cost function will be a by-product of choosing the best input-output technique over the planning interval. Certain interesting results follow from this construction of the problem.

Following Modigliani and Hohn, we assume a finite planning horizon of T years, and take as our criterion the minimization of undiscounted cumulative cost over this period. Suppose the amount of product inventory held at time t is $H(t)$. Then if $Y(t)$ is the unknown cumula-

[9] See Morin [4], and Arrow, Karlin, and Scarf [1], pp. 61–69, for continuous formulations of the Modigliani-Hohn model.

tive production up to t, and $S(t)$ is the given cumulative sales up to t, we have $H(t) = H(0) + Y(t) - S(t)$. That is, inventory at t is initial inventory plus the excess of cumulative production over cumulative sales at t. We assume one current and one capital equipment input to the process. We let k be the constant marginal cost of holding a unit of inventory per unit of time.

Cumulative cost over the horizon can now be written

$$C = \int_0^T (w_1 x_1(t) + k[H(0) + Y(t) - S(t)])dt + W_2 X_2 . \tag{5.6}$$

The problem is to choose $x_1(t)$, X_2, and $Y(t)$ so as to minimize (5.6), subject to the production function constraint

$$f[x_1(t), X_2] - \dot{Y}(t) = 0 . \tag{5.7}$$

Also, we have the important nonnegativity condition on inventory stocks,

$$H(t) = H(0) + Y(t) - S(t) \geq 0 , \tag{5.8}$$

and the boundary condition requiring no inventory at the end of the planning period,

$$Y(T) = \int_0^T f[x_1(t), X_2]dt = S(T) - H(0) . \tag{5.9}$$

Necessary conditions for such a minimum are

$$w_1 - \lambda f_1[x_1(t), X_2] = 0 , \tag{5.10}$$

$$\int_0^T -\lambda f_2[x_1(t), X_2]dt + W_2 = 0 , \tag{5.11}$$

and

$$k - \frac{d\lambda}{dt} \geq 0 . \tag{5.12}$$

It is assumed that equalities hold in (5.10) and (5.11), since inequalities here are not of interest if some of both inputs are required to produce any positive output. λ is the short-run or operating marginal cost function, assumed given by Modigliani and Hohn.

The new center of interest is the Euler condition (5.12). Integrating this condition from t to $t + \theta$, where $0 \leq t + \theta \leq T$, gives

$$\int_t^{t+\theta} k dt - \lambda(t + \theta) + \lambda(t) \geq 0 ,$$

or

$$k\theta + \lambda(t) \geq \lambda(t + \theta) . \tag{5.13}$$

The left side of (5.13) is the marginal cost of producing a unit at time t and holding it in inventory until time $t+\theta$. The right side of (5.13) is the marginal cost of producing a unit at $t+\theta$. Hence, (5.13) states that the marginal cost of producing a unit at any given time must not exceed the cost of producing it at an earlier time and holding it until the given time.

If the inequality in (5.13) holds over some interval $t_1 \leq t+\theta \leq t_2$, then in that interval it costs more to produce a unit at t and hold it until any subsequent time, than to produce the unit at the later time. Therefore, at any t in this interval, we should make $Y(t)$ as small as possible subject to the condition (5.8). If inventories are zero at the opening of the interval, i.e., $H(t_1) = H(0) + Y(t_1) - S(t_1) = 0$, then the smallest we can make $Y(t)$ is $Y(t) = S(t) - H(0)$, and current sales are satisfied out of current production in the interval. If inventories are positive at the opening of the interval, i.e., $H(t_1) = H(0) + Y(t_1) - S(t_1) > 0$, then we make $Y(t) = Y(t_1)$, scheduling a zero level of current production, until such time as the opening inventory is exhausted. Thereafter, the previous solution applies, with current sales met out of current production. Of course, over any interval for which current production equals current sales requirements, i.e., $\dot{Y}(t) = \dot{S}(t)$, the inventory part of the model disappears. The solutions for $x_1^0(t)$, X_2^0, and $\lambda^0(t)$ are then obtained by simultaneously satisfying $\int_0^T \frac{w_1 f_2}{f_1} dt = W_2$, $\dot{S}(t) = f[x_1(t), X_2]$, and $\lambda(t) = \frac{w_1}{f_1[x_1(t), X_2]}$.

If the equality holds in (5.12) over some interval, the solutions $x_1^0(t)$, X_2^0, $\lambda^0(t)$, and $Y^0(t)$ are obtained by simultaneously satisfying (5.7)–(5.12). However, in obtaining these solutions, the latter conditions are separable in a most interesting way. Suppose, for simplicity, we assume the equality in (5.12) to hold over the entire planning horizon. By integrating (5.12), and evaluating the integration constant at $t=0$ and $t=T$, we get $kt + \lambda(0) - \lambda(t) = k(t-T) + \lambda(T) - \lambda(t) = 0$, which defines marginal *production cost* as a (linear) function of time. Hence, insofar as the short-run marginal cost function depends upon time, this relationship is determined by the differential equation (5.12) *independently of sales requirements, $S(t)$, and the production function.* This is in sharp contrast to the no-inventory case in which $x_1^0(t)$ and X_2^0 are jointly determined by the dynamic marginal-rate-of-substitution condition $\int_0^T (w_1 f_2/f_1)dt = W_2$ and the technological constraint $\dot{Y}(t) = \dot{S}(t) = f[x_1(t), X_2]$. Only after $x_1^0(t)$ and X_2^0 are so determined do we obtain the equilibrium dynamic marginal cost function from $\lambda(t) = w_1/f[x_1(t), X_2]$. In the no-inventory case the dependence of marginal cost on time is

obtained only after we know the optimal path of current input consumption, $x_1(t)$. But the moment we introduce inventories, fix sales requirements, and allow output to be a decision variable, the variation in marginal production costs with time is no longer due to the marginal productivity conditions and the production function constraint. With inventories, production planning is constrained by the requirement that the marginal cost of producing a unit at any time be equal to the marginal cost of producing a unit at any earlier time and holding it to the given time. The consumption of current input over time, and, therefore, the marginal cost function over time, must conform to this condition. The structure of optimal production policy is altered thereby in a manner which may not be *prima facie* obvious.

Assuming the equality holds in (5.12), we can write the simultaneous conditions on $x_1(t)$, X_2, $\lambda(t)$, $Y(t)$, and $x_1(T)$, as follows:

$$k(t-T)+\frac{w_1}{f_1[x_1(T),X_2]}-\frac{w_1}{f_1[x_1(t),X_2]}=0\,, \tag{5.14}$$

$$\int_0^T \frac{w_1 f_2[x_1(t),X_2]}{f_1[x_1(t),X_2]}dt=W_2\,, \tag{5.15}$$

$$Y(T)=\int_0^T f[x_1(t),X_2]dt=S(T)-H(0)\,, \tag{5.16}$$

$$\lambda(t)=k(t-T)+\frac{w_1}{f_1[x_1(T),X_2]}\,, \tag{5.17}$$

$$Y(t)=\int_0^t f[x_1(t),X_2]dt\,. \tag{5.18}$$

Equations (5.14)–(5.16) determine jointly the variables $x_1^0(T)$ and X_2^0, and the extremal $x_1^0(t)$. Using the solutions X_2^0 and $x_1^0(T)$, we obtain $\lambda^0(t)$ from (5.17) as a function of t and all the parameters of (5.14)–(5.16). Finally, using $x_1^0(t)$ and X_2^0, we obtain $Y^0(t)$ from (5.18). The same procedure applies for any subinterval over which the equality holds in (5.12).

In comparing these results with those of the Modigliani-Hohn model, the important thing to observe is the different role played by (5.12). In models of the Modigliani-Hohn type, the marginal cost function $\lambda=\lambda[\dot{Y}(t)]$ is a given function of current output *independent* of inventory and production planning considerations. Under such conditions (5.12) places a constraint on $\dot{Y}(t)$ only, whereas in our model, in which λ is dependent upon choice of technique, the inventory condition (5.12) places a constraint on both λ and $\dot{Y}(t)$.

Purdue University, U.S.A.

REFERENCES

[1] ARROW, K., S. KARLIN, AND H. SCARF: *Studies in the Mathematical Theory of Inventory and Production*. Stanford: Stanford University Press, 1958.

[2] FOX, CHARLES: *An Introduction to the Calculus of Variations*. London: Oxford University Press, 1950.

[3] MODIGLIANI, F. AND F. HOHN: "Production Planning Over Time and the Nature of the Expectation and Planning Horizon," *Econometrica*, Vol. XXIII (January, 1955), pp. 46-66.

[4] MORIN, FRANÇOIS: "Note on an Inventory Problem," *Econometrica*, Vol. XXIII (October, 1955), pp. 447-450.

[5] SMITH, VERNON L.: "The Theory of Investment and Production," *Quarterly Journal of Economics*, Vol. LXXIII (February, 1959), pp. 61-87.

[6] ———: "Economic Replacement Policies: An Evaluation," *Management Science*, Vol. IV (October, 1957), pp. 20-37.

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3, September, 1960

SIR ROY F. HARROD ON INTERNAL AND EXTERNAL BALANCE*

BY MASAHIRO TATEMOTO[1]

1. THE DISCUSSION of possible conflicts between the policy requirements of full employment (without inflation) and those of equilibrium balance of payments is an important application of the Keynesian theory of employment. This problem has been one of the most important issues in postwar international economics, and many writers have been concerned with it.[2] Among them, Sir Roy F. Harrod, in the new edition of his book,[3] devoted an ingenious chapter to "the theory of balance of payments by considering what should be done in order not only to get an equal balance of external payments, but also at the same time to maintain full employment at home." In this short note we shall put his theory in a diagrammatic form and make the point more clearly.

He defines the employment equilibrium or internal balance as $Y^d = Y^s$, where Y^d stands for aggregate demand and Y^s for the "supply potential" or "full employment income." In the case where aggregate demand is less than the supply potential, realized income, Y, is equal to aggregate demand, i.e., $Y^d = Y < Y^s$, and there is unemployment. If aggregate demand exceeds the supply potential, realized income equals the supply potential, i.e., $Y^d > Y = Y^s$, where the inflationary gap[4] is measured by $Y^d - Y^s$.

* Manuscript received October 16, 1959.

[1] The writer has profited greatly from many helpful suggestions and comments made by Professors Michio Morishima of Osaka University, Anthony M. Tang of Vanderbilt University, and Harry G. Johnson of the University of Chicago.

[2] It was the late Professor Ragnar Nurkse who first pointed out the importance of this problem. See R. Nurkse, "Domestic and International Equilibrium," in S. E. Harris, ed., *The New Economics, Keynes' Influence on Theory and Public Policy* (London and New York: D. Dobson, Ltd. and A. Knopf, Inc., 1947), pp. 264 ff. Professor J. E. Meade also discussed in detail the "conflicts between internal and external balance" in his *The Balance of Payments* (London and New York: Oxford University Press, 1951), Chapter X, pp. 114 ff.

[3] Harrod, R. F., *International Economics*, Fourth edition, Chapter VII, (Cambridge Economic Handbook, Nisbet and Cambridge University Press, 1957).

[4] This excess of aggregate demand over the supply potential is called "the quantum of unfulfilled orders." If prices rise so as to reduce demand to the supply potential, the inflationary gap will be cleared at once. The classical school is concerned with this case. But it is probable that prices will not rise so much and that the inflationary gap takes the form of unfilled orders or waiting list. This is the Harrod case.

External balance is defined as the equality of exports and imports, $X = M$, where X and M denote the exports and the imports respectively. In the event that $X \gtrless M$, there exists "external disequilibrium or imbalance."[5]

If $Y^d = Y^s$ and $X = M$, the economy is in equilibrium in both respects. But there may be an employment disequilibrium, $Y^d \gtrless Y^s$, while there is an equal external balance, $X = M$. Or there may be an unequal external balance, $X \gtrless M$, even though the internal economy is balanced, $Y^d = Y^s$. Thus, Harrod sets out the following four combinations of imbalance and proposes corrective remedies for them.[6]

Case	Situations		Remedies
I	$Y^d < Y^s$ and $X > M$	(unemployment and external surplus)	Reflate
II	$Y^d > Y^s$ and $X < M$	(inflation and external deficit)	Disinflate
III	$Y^d < Y^s$ and $X < M$	(unemployment and external deficit)	Reduce factor rewards or devalue currency
IV	$Y^d > Y^s$ and $X > M$	(inflation and external surplus)	Increase factor rewards or revalue currency upward

2. It may be of some interest to reformulate Harrod's verbal discussions in terms of the ordinary 45° line diagram of national income determination. Instead of Harrod's concept of national income, defined as the total sum of the expenditure on home-produced consumer goods, home-produced investment goods, and exports (excluding import contents), we shall here use a more popular definition of national expenditure, i.e.,

$$Y = C + I + G + X - M,$$

where C, I, and G represent consumption, investment, and government expenditure respectively. Figure 1 is a slightly modified 45° line diagram of income determination. The level of national income is measured along the horizontal axis; exports, X, and imports, M, are measured downward along the vertical axis, and the aggregate demand, $C + I + G + X - M$, upward. Suppose the aggregate demand is represented by the line aa'; then the income, Y, corresponding to the intersection A of the line aa' and the 45° line stands for realized in-

[5] It is assumed that there are no gifts and no spontaneous movement of capital. Harrod distinguishes the imports of consumer goods from those of investment goods, but here we use the aggregate category, "the imports", for simplicity.

[6] Harrod, *op. cit.*, pp. 143–152.

come at which total demand equates total supply. The supply potential or full employment income is represented in our diagram by the income level Y^s, and there is unemployment, which is represented by the distance YY^s. Yet the income level Y^e which warrants the external balance (given by the intersection of the lines X and M) lies in our diagram to the left of Y, and there is an external deficit represented by the vertical distance EF. Thus, Figure 1 is the graphical representation of Harrod's Case III, where unemployment and external deficit are jointly present. Suppose some expansionist measures (for instance, fiscal policy) to increase domestic expenditures (I or G) are taken; then, in our diagram, aggregate demand will be shifted upward to bb', and the full employment level Y^s will be attained. This will, however, lead to the widening of the gap between imports and exports (viz., from EF to $E'F'$). On the contrary, deflationary measures will succeed in recovering the external balance by shifting the aggregate demand downward to cc', but it will lead to an increase in unemployment (viz., from YY^s to Y^eY^s). This is Harrod's "doubly misfortunate" case in the sense that expansion aggravates the external

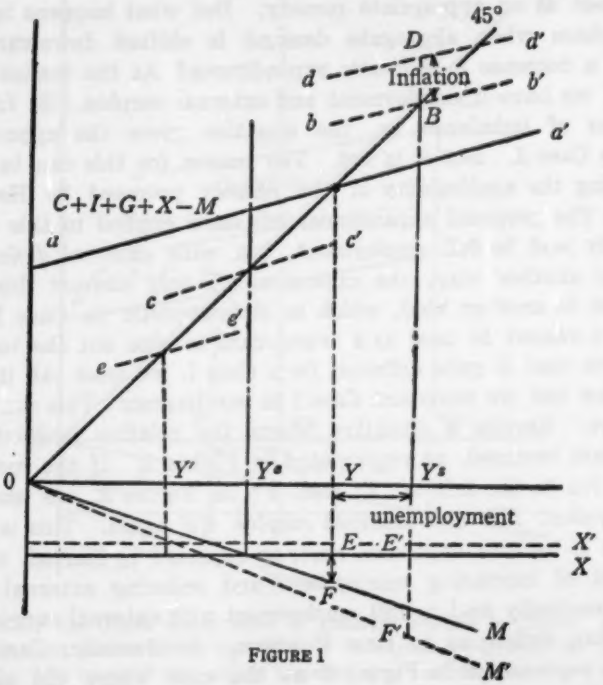

FIGURE 1

balance, while contraction disturbs the internal balance. Therefore, the increase in aggregate demand should not be effected by increasing domestic expenditures I or G, but by increasing X or reducing M. The latter will have the effect of shifting Y^e to the right. The appropriate remedy will, according to Harrod, be a reduction (either selective or general) of domestic factor rewards or a devaluation of the currency. This will, subject to the exchange stability condition, lead to an increase in X and a decrease in M and cause the required shift to Y^e.

3. The above diagram can readily be modified to represent the other cases. If the aggregate demand is shifted upward to dd' through an increase in *domestic* expenditures, we will have the inflationary gap represented by the vertical distance DB. This inflation will cause an increase in imports and a decrease in exports by making home-made goods more expensive than foreign goods. The line M will shift downwards and the line X upwards, as represented by the dotted lines in Figure 1. These two effects will lead to an aggravated external imbalance. This is Harrod's Case II, for which disinflation is proposed as an appropriate remedy. But what happens in the opposite phase when aggregate demand is shifted downward to ee' through a decrease in domestic expenditures? At the realized income level Y' we have unemployment and external surplus. So far as the symptoms of imbalance go, the situation gives the appearance of Harrod's Case I. But it is *not*. The reason for this can be seen by considering the applicability of the remedy proposed by Harrod for Case I. The proposed expansionist measures applied to this case will eventually lead to full employment, but with external *deficit*. Putting it in another way, the expansion will only convert this type of imbalance to another kind, which is characteristic to Case III; thus expansion cannot be used as a trump card to wipe out the imbalance. The above case is quite different from Case I; we shall call it Case I'.

But how can we represent Case I in our diagram? This can be done as follows. Assume a situation where the relative positions of Y^e and Y^e are reversed, as represented in Figure 2. If the realized income Y lies to the left of Y^e and Y^e in Figure 2, we shall have unemployment YY^e and external surplus EF again. This is nothing but Case I, because expansion here, as expected by Harrod, will have the effect of increasing employment and reducing external surplus. It will eventually lead to full employment with external *surplus* $E'F'$, rather than deficit as in Case I' above. Incidentally, Case IV can easily be represented in Figure 2 as the case where the aggregate

demand exceeds the full employment income $OY^f(=BY^f)$ by BC.

According to the opinions of government economists here, an actual example of Case I′ is provided by the Japanese experience in the 1957–58 recession, which was characterized by under-utilization of capital equipment[7] and external surplus. It may be considered that Y^e lay to the left of Y^f in this Japanese situation. Expansion was out of the question because it came shortly after the government adopted deflationary measures to correct the big external deficit (which appeared in the first quarter of 1957, when the economy approached the full capacity ceiling). A similar situation is likely to occur in an economy which depends largely on imported raw materials.

Cases I and I′ above are twin brothers but have quite different dispositions of their own. As the reader observes, this difference comes from the different relative positions of Y^e and Y^s, reflecting the structural basis of an economy. The policy of expansion should not be applied to Case I′, since it will in the process convert Case

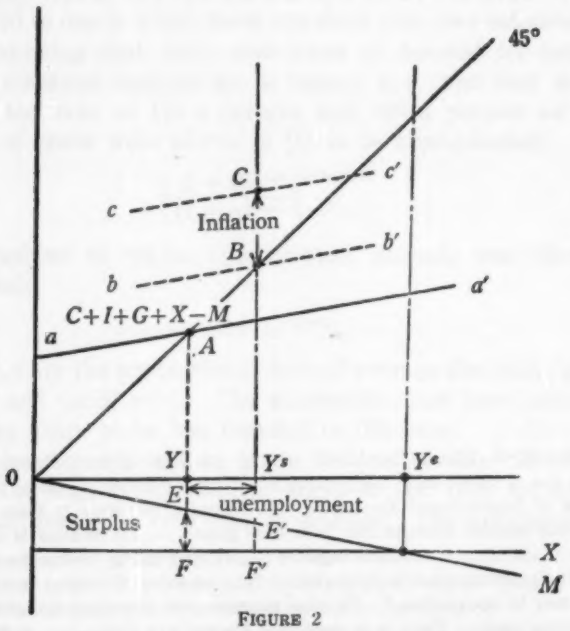

FIGURE 2

[7] In a country like Japan where labor is abundant relative to capital, the supply potential Y^s is determined by the full capacity of capital equipment, but not by full employment. The excess of the supply potential over realized income is measured by the under-utilization of capital equipment, but not by the unemployment of labor.

I′ to Case III. Appropriate remedies for Case I′ would then be a combination of the two remedies proposed for Case I and Case III. These are a policy of expansion until Y^e is reached and, beyond Y^e, a policy of devaluation or factor reward reduction. The author believes that it is very important to recognize the practical implications of the relative positions of Y' and Y^e.[8]

Osaka University, Japan

[8] Professor H. G. Johnson has kindly pointed out that Harrod distinguished between "the main line of cure" and "the possible trimming" operation in an article preceding the new edition of *International Economics*. For example, referring to Case I, Harrod suggests: "If the internal stimulus has to be very great. . ., the balance of trade may shift unfavorably, and if . . . it becomes negative . . ., it may in due course become needful to trim the foreign exchange rate downwards. If . . . an external surplus remains, an upward trimming may be appropriate." He also suggests the trimming operations for each of the other three cases. Thus, it is clear that Harrod was quite aware of the possiblity of Case I′. However, in his *International Economics*, there is no distinction between a major and a minor remedy. See Harrod, R. F., "Currency Appreciation as An Anti-Inflationary Device: Comment," *Quarterly Journal of Economics*, Vol. LXVI, No. 1, 1952, pp. 102–116.

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3, September, 1960

TARIFFS AND TRADE IN A TWO-GOOD WORLD*

BY W. M. GORMAN[1]

1. INTRODUCTION

THIS SHORT NOTE is a sequel to "The Effect of Tariffs on the Level and Terms of Trade" [1], some of the results of which are extended to the case in which there are more than two countries.

The elasticity of demand for the exports of a single country is discussed in Section 2. The argument relies heavily on the assumption that there are only two goods, so the conclusion that this elasticity is likely to be high need not be taken too seriously. Both argument and conclusions, however, are so simple that they seem worth presenting; they may, after all, form the basis for more useful work.[2]

The general results proved for a two-country, two-good world in [1] are extended to one in which there are more than two countries in Section 2. Assuming that their elasticities of demand for import were j, k, their marginal propensities to import ρ, σ, and that they levied tariffs at the rate of 100 c percent and 100 d percent *ad valorem*, the terms of trade were shown in [1] to be approximately

$$\left\{\frac{(1+d)^{k-\sigma}}{(1+c)^{j-\rho}}\right\}^{1/(j+k-1)},$$

while the volume of trade, appropriately defined, was shown to be approximately

$$(1+t)^{-(m-\theta)},$$

where m, θ, t are the appropriately defined average elasticity, propensity to import, and tariff level.[3] The assumption that there are only two goods seems likely to be less harmful in this case.

This seems an appropriate place for me to say that I wish to withdraw the suggestion, strongly urged in [1], that we may expect trade elasticities to be very high in practice. The argument for that

* Manuscript received January 4, 1960.

[1] I am grateful to Mr. Esra Bennathan, Dr. Frank Hahn, and D. John Wise for their comments. None of them should be blamed for my inadequacies.

[2] Besides, they are needed in Section 2. At least one reader found them very surprising, so, perhaps, others may too.

[3] Units being chosen so that both the terms and volume of trade are unity in the absence of tariffs.

proposition depended critically on the assumption that the country in question produced a perfect substitute for the good it imported. Professor L. R. Klein has persuaded me that this is unreasonable.[4]

A fellow economist who read this paper was under the impression that its conclusions were valid only if each country's share of the market were fixed. The paper would not be very interesting if this were so, for it would imply that the various offer curves on each side differed only in scale and therefore had the same elasticity. I differentiate first and *then* put the observed market share into my formulae, so that they can be applied to the actual world where market shares are changeable.[5]

2. THE ELASTICITY OF DEMAND FOR A SINGLE COUNTRY'S EXPORTS

Consider a world consisting of N countries trading in two goods X and Y. A_i exports x_i units of X in return for y_i units of Y ($i = 1, 2, \cdots, n$), and B_r exports y_r units of Y in return for x_r of X ($r = n+1, n+2, \cdots, N$).[6] A_i's elasticity of demand for imports is j_i; B_r's is k_r.

Assume that each country's trade is balanced and that the world price of X in terms of Y is P.[7] Then,

$$(1) \qquad y_i = Px_i; \qquad y_r = Px_r; \qquad \text{and } y = Px,$$

where

$$(2) \qquad x = \sum x_i = \sum x_r; \qquad y = \sum y_i = \sum y_r$$

stand for the volume of world trade in X and Y respectively. We can therefore define

$$(3) \qquad \lambda_i = x_i/x = y_i/y; \qquad \mu_r = x_r/x = y_r/y$$

unambiguously as A_i's and B_r's share in the world market respectively.

We shall now seek the elasticity (with fixed tariffs) of *demand* K_i for A_i's exports. To do so, we displace A_i's offer curve[8] and see what happens to the net foreign demand $\bar{x}_i = \sum x_r - \sum_{h \neq i} x_h$, for A_i's exports.

[4] In combination with Sir Donald MacDougall's fine study of British and American exports [2], which I have recently reread.

[5] Of course the reader may find some of the *other* assumptions unrealistic.

[6] Throughout this paper h and i will vary over the A's, r over the B's.

[7] Indeed we make all the usual assumptions of optimum tariff theory—except, of course, that the tariffs need not be optimal and that there may be more than two countries.

[8] Possibly by allowing A_i to levy a new tariff, possibly otherwise. Note that we are estimating the foreign *demand* for A_i's exports and must therefore displace its own offer curve.

By definition,[9]

$$(4) \qquad k_r = -P\partial x_r/x_r\partial P; \qquad j_i = P\partial y_i/y_i\partial P,$$

therefore,

$$(5) \qquad P\partial x_r/\partial P = -k_r x_r = -\mu_r k_r x \qquad \text{by (3)},$$

and, by (1), (3), and (4),

$$(6) \qquad P\partial x_h/\partial P = \partial y_h/\partial P - x_h = \lambda_h(j_h - 1)x.$$

(6) gives the change in the *supply* of A_h's exports when their price increases by dP.

The net increase in the demand for A_i's exports is

$$\partial \bar{x}_i/\partial P = \sum \partial x_r/\partial P - \sum_{h \neq i} \partial x_h/\partial P,$$

therefore,

$$(7) \qquad -P\partial \bar{x}_i/\partial P = [j + k - 1 - \lambda_i(j_i - 1)]x,$$

where $j = \sum \lambda_i j_i$ and $k = \sum \mu_r k_r$ are the average import elasticities of demand for Y and X respectively.[10]

Initially, $\bar{x}_i = x_i = \bar{\lambda}_i x$, so that

$$(8) \qquad K_i = (j + k - 1)/\lambda_i - (j_i - 1).$$

The elasticity of demand for B_r's exports is similarly

$$(9) \qquad J_r = (j + k - 1)/\mu_r - (k_r - 1).$$

Unless A_i dominates the world market, the leading terms in (8) and (9) will be the more important, so that[11]

$$(10) \qquad K_i \doteq (j + k - 1)/\lambda_i \quad \text{and} \quad J_r \doteq (j + k - 1)/\mu_r$$

and, in particular,

$$K_i \doteq J_r \quad \text{if} \quad \lambda_i \doteq \mu_r.$$

Hence, the elasticity of demand for a country's exports depends mainly on its share of the market and is largely independent of the good which it exports. It seems likely to be high, too. If $j_i = j =$

[9] $j_i = -\partial \log y_i/\partial \log (1/P) = \partial \log y_i/\partial \log P$. The partial ∂'s are used to denote the fact that *tariffs* are held constant in the other countries. The demand and supply relations are given in (11) below.

[10] Appropriately weighted.

[11] If all the j's and k's are greater than one:

$$j - 1 - \lambda_i(j_i - 1) = \sum_{h \neq i} \lambda_h(j_h - 1) \geqq 0,$$

so that $K_i \geqq k/\lambda_i$ and $J_r \geqq j/\mu_r$.

$k = 2$, for instance, and A_i has twenty percent of the world market, $K_i = 14$.[12]

What happens if other countries change their tariffs as well?

Suppose A_i levies a tariff of 100 t_i percent *ad valorem*, $(i=1, 2, \cdots, n)$, and B_r one of 100 t_r percent, $(r = n + 1, \cdots, N)$. We can then write

$$x_i = F_i(P, t_i), \qquad x_r = G_r(P, t_r) \tag{11}$$

and find

$$Pdx_i = (j_i - 1)x_i dP + Px_i' dt_i, \tag{12}$$

$$Pdx_r = -k_r x_r dP + Px_r' dt_r, \tag{13}$$

where

$$x_i' = \partial x_i/\partial t_i, \qquad x_r' = \partial x_r/\partial t_r, \tag{14}$$

with the aid of (4).

Hence, by (2) and (7),

$$(j + k - 1)xdP = P(\sum x_r' dt_r - \sum x_s' dt_s). \tag{15}$$

Substituting into (12) we find the total elasticity of demand for A_i's exports,

$$\bar{K}_i = [\theta_i(j + k - 1)/\lambda_i] - (j_i - 1), \tag{16}$$

where

$$\theta_i = x_i' dt_i/(\sum x_s' dt_s - \sum x_r' dt_r). \tag{17}$$

If A_i is the only country to change its tariff, $\theta_i = 1$, and (16) is the same as (8)—as, of course, it should be.

If A's competitors—the other A's—raise their tariffs to the same extent, but the B's do nothing, and if the A's are similar to each other in the sense that x_s'/x_s is the same for all of them,[13] $\theta_i = \lambda_i$, and $\bar{K}_i = j - j_i + k$. Finally, if all this is true and $j_i = j$, so that A_i is truly an "average" country[14]—except, perhaps in size:

$$\bar{K}_i = k\,. \tag{18}$$

In other words, if the A's are sufficiently similar[15] and act in a sufficiently similar manner, each of them is effectively in a monopoly position, as one would expect.

In general, however, we know nothing at all about $\bar{K}_i$, not even

[12] Please remember that these results depend on the assumption that the consumer is indifferent between units of, for example, X made in different countries.

[13] Quite a reasonable assumption, since it refers only to *proportional* shifts.

[14] Alternatively, one might argue that (18) were true "on the average" in some sense.

[15] They may, of course, be different sizes.

its sign. The implications for the theory of optimum tariffs are obvious. On the other hand, we must remember that these results have been proved under the assumption that the goods exported by any one of the A's are perfect substitutes for those exported by the other.

3. TARIFFS AND TRADE

In this section we shall show that the results proved in Sections II and III of [1] carry over to a many country world, at least as first order approximations.

It will be convenient to consider first the effects of tariffs on the terms of trade P,[16] and to normalize our data so that, under free trade,

$$x = y = P = 1. \tag{1}$$

(2.15)[17] implies that

$$d \log P = (\sum x_r' dt_r - \sum x_i' dt_i)/(j + k - 1). \tag{2}$$

To convert this into a useful form, we must express x_i' and x_r' in terms of the marginal propensities ρ_i and σ_r to import into A_i and B_r respectively. In order to do so, we first define

$$Q_i = P/(1 + t_i) \tag{3}$$

as the domestic price of X in terms of Y in A_i, and

$$Z_i = y_i - Q_i x_i = Q_i x_i t_i, \quad \text{since } y_i = P x_i, \tag{4}$$

as A_i's exogenous income. (2.12) and (3) imply that

$$dy_i = P dx_i + x_i dP = (j_i Q_i x_i + x_i') dt_i \tag{5}$$

when Q_i is held constant. At the free trade equilibrium,

$$\rho_i = (\partial y_i/\partial Z_i)_{Q_i \text{ constant}} = j_i + x_i'/x_i. \tag{6}$$

Similarly,

$$\sigma_r = k_r + x_r'/x_r.^{18} \tag{7}$$

Substituting into (2), we find tha

$$\log P \doteq \frac{(j - \rho)c - (k - \sigma)d}{j + k - 1} \tag{8}$$

to the first order for "small" tariffs, where

[16] Strictly speaking, P is the terms of trade of the B's. It corresponds to $1/P$ in [1].
[17] We refer thus to equations in other sections.
[18] This can easily be verified.

$$j = \sum \lambda_i j_i; \qquad k = \sum \mu_r k_r;$$
$$\rho = \sum \lambda_i \rho_i; \qquad \sigma = \sum \mu_r \sigma_r, \tag{9}$$
$$c = \frac{\sum \lambda_i (j_i - \rho_i) t_i}{\sum \lambda_i (j_i - \rho_i)}; \qquad d = \frac{\sum \mu_r (k_r - \sigma_r) t_r}{\sum \mu_r (k_r - \sigma_r)}$$

are appropriately weighted averages in each case. This is the result which we found in [1] for trade between two countries A and B, whose marginal propensities to import were ρ and σ, whose elasticities of demand for imports were j and k, and which levied tariffs of $100c$ and $100d$ percent *ad valorem* respectively.[19] It would therefore seem that our two groups of countries can be aggregated to form two imaginary superstates whose elasticities, marginal propensities to import, and tariffs are "averages" of those for their component economies. This sample pattern of aggregation is possible only because there is no trade between the members of the same group, and because the tariffs are sufficiently small to ensure that no exporter of X becomes an exporter of Y, or *vice versa*.

We can deduce from (8), as we did in [1], that tariffs are likely to have a negligible effect on the terms of trade.[20]

Let us next consider the volume of trade, which we define as being $I = \sqrt{xy}$. Under our normalization, $I = 1$ under free trade.

Eliminating x and P from (2.3), (2.12) and (2.13), and using (8) and (9), we find that

$$d \log I \doteq d \log x + 0.5\, d \log P,$$

therefore,

$$\log I \doteq -(m - \theta)t, \tag{10}$$

where m, θ, and t stand respectively for the "average" elasticity of demand in the export market, the "average" marginal propensity to import, and the "average" tariff, as defined in [1]. The averaging process is here carried out over all N countries.

This again corresponds to the main result in Section II of [1]. The precise definitions of m, θ, and t are given in Appendix B to that paper. The quality of the approximation is only slightly impaired if we take

$$\begin{aligned} m - 0.5 &= \text{the harmonic mean of } j - 0.5 \text{ and } k - 0.5, \\ \theta &= \text{the arithmetic mean of } \rho \text{ and } \sigma, \\ 1 + t &= \text{the geometric mean of } 1 + c \text{ and } 1 + d. \end{aligned} \tag{11}$$

[19] Of course, this is merely a first order approximation here. The result in [1] looks rather different, but is the same to the first order.

[20] See the example in [1].

The main theoretical results proved for a two-country, two-good world in [1] are seen to hold for a many-country, two-good world, even if in a somewhat attenuated form.[21] The main effect of the existence of more than two goods is, presumably, the possibility of multilateral trade,[22] which is probably more vulnerable to tariffs than bilateral, because it has to surmount several barriers. These formulae may well understate the effect of tariffs in a many-good world, but it is not clear by how much. There is some discussion of this point in [1].

University of Birmingham, England

REFERENCES

[1] GORMAN, W. M., "The Effects of Tariffs on the Level and Terms of Trade," *Journal of Political Economy*, Vol. LXVII, 1959, pp. 246–65.

[2] MACDOUGALL, G.D.A., "British and American Exports; a Study Suggested by the Theory of Comparative Costs," *Economic Journal*, Vol. LXI, 1951, pp. 697–724, and Vol. LXII, 1952, pp. 487–521.

[21] Though in an attenuated form: (i) these are approximations for small tariffs, (ii) there are no analogues of Propositions I and II of [1].

[22] Of essentially multilateral trade, that is, in which A sends X to B, B sends Y to C, and C sends Z to A, for instance.

INTERNATIONAL ECONOMIC REVIEW
Vol. 1, No. 3, September, 1960

THE LONG-RUN DEVELOPMENT OF NATIONAL INCOME IN GERMANY: A REVIEW ARTICLE*

By Bruno Dietrichs and Reimut Jochimsen

1. INTRODUCTION

Research in long-run economic development is today one of the most important objectives in economics. Among the foundations for these investigations are the works of J. M. Keynes, who made the determination and fluctuation of national income his primary goal. Since then, starting from the determination of national income and employment in a closed system, models have been devised which make possible an analysis of the process of economic growth. These models analyze many relevant variables such as income, employment, investment, and consumption as they change over time.

With his theory of national income Keynes demonstrated that a whole economy may be represented by certain macro-economic aggregates, and that general economic equilibrium can be described by a few strategic variables. In this way, rather independent statistics of income, savings, and investment may be used to verify statistically economic theory. Keynes' theoretical starting point, however, is the concentration of an infinite number of elementary economic units into "aggregates relevant for the functioning of the economy."[1] His objective is thus a short-run analysis of the economic process and not an examination of long-term growth. Keynes recognized these limitations and mentioned them explicitly in many places.[2] Similarly, many extensions of his work by his successors are limited to typical short-run problems.

Many attempts have recently been made to set up theoretical models which will study long-run aspects of growth, utilizing the indispensable instruments of Keynesian theory. The analysis of short-run fluctuations around the trend of "equilibrium" is replaced by study of the trend itself. There had been attempts by classical economists to understand long-run economic development, even if only in

* Manuscript received February 25, 1960.

[1] Erich Schneider, *Einführung in die Wirtschaftstheorie*, Vol. I (Tübingen: 3rd ed., 1955), p. 50.

[2] Gottfried Bombach, "Zur Theorie des wirtschaftlichen Wachstums," *Weltwirtschaftliches Archiv*, Bd. 70 (1953, I), p. 110.

terms of secular stagnation; now, however, there is a renascence from the type of price and market theory—microeconomic, polypolistic, without regard to time—which for so long dominated economic thinking. Harrod, Domar, Baumol, Hicks, and Bombach have devised important long-run models; in particular, Harrod and Hicks have worked on a theoretical synthesis of short- and long-run analysis.

Although current investigations of economic growth are still predominantly theoretical, it has proved possible in some cases to use for verification statistical material originally prepared for the typical Keynesian short-run analysis. The study of changes in the level of national income over time has long been of primary importance in economic research, especially in growth analysis; empirical testing of models of economic growth should therefore start with the statistical determination of the nominal value of national income over a continuous long period.

2. ORIGINS OF NATIONAL INCOME STATISTICS IN GERMANY

In the case of Germany, quite extensive statistical material is available over a long period which, if properly prepared, could throw light on national income. This situation appears to be rather unique, since in other highly industrialized economies the relevant statistics do not go as far back into the last century as do the German ones. Now at last the scientific and methodological objectives of the German Historical School (which originally dominated German economic thought and later was almost entirely superceded by the Mathematical School) may be justified, ironically enough, by the very theoreticians who earlier fought it. The collection and publication of very detailed data about many different time series such as population changes, income tax payments, etc. are the results to a very great extent of the emphasis which the Historical School placed on inductive methods. The importance of these series should not be overlooked, even though they were often too extensive and badly arranged and, in any case, became of real significance only after evaluation by deductive methods. The representatives of the German Historical School—like true children of their time—were to a large degree simply following a trend current in all sciences, natural, social, or humanistic, by attempting to push forward the boundaries of knowledge through the encyclopedic collection of facts.

3. THE HOFFMANN-MÜLLER STUDY OF GERMAN NATIONAL INCOME

For these reasons, it has been possible to utilize the German figures

in order to verify certain aspects of modern growth theory. Professor Walther G. Hoffmann of the University of Münster and Professor J. Heinz Müller of the University of Freiburg, together with their staff of assistants and research workers, have succeeded in constructing for the first time national income figures for Germany over a period of more than one hundred years, i.e., from 1851 onwards. Moreover, these series are presented in a form which permits comparisons.[3] These statistical results may make possible a threefold study of the process of growth of the German economy: first, the inductive explanation of growth; second, the testing of dynamic theories; and third, evaluation of future opportunities for development. They may also provide a basis for the determination of structural changes in the German economy from 1851 to 1957 by using cross-sectional stocktaking.

Understandably, a collection of statistical series going this far back into the nineteenth century presents extraordinary problems and difficulties. The authors must, however, be commended for taking this risk, for the results that have been achieved fully justify the undertaking. The book is limited to a reproduction of statistical material in the past hundred years; any theoretical interpretation of the results must be left to further investigators, since a structural analysis of the sources of national income is not given.[4] The greater part of the Hoffmann-Müller volume is devoted to an extensive examination of the nominal value of national income. Estimation of a parallel time series of real national income is hindered by the lack of price indexes proper for this type of macroeconomic investigation. Any price indexes available so far back usually omit one or another important component, the inclusion of which is essential for proper deflation of the nominal values.

Let us take, for example, price indexes before 1913. Price changes in essential foodstuffs, which frequently move contrary to the cost of living index, are overrepresented; in addition, the available index of wholesale prices does not sufficiently reflect the development of purchasing power. Taking these difficulties into account, the authors

[3] W. G. Hoffmann, J. H. Müller, with the collaboration of Heinz König and Klaus Döring, Jochen Dohmen, Rolf Goldschmidt, Franz Grumbach, Alexander Hölling, Arnold Hülden, *Das deutsche Volkseinkommen, 1851–1957* (Tübingen: J. C. B. Mohr (Paul Siebeck), 1959).

[4] Estimates for the generation of national product and the spending of national income are still lacking. The first problem will be extremely difficult to solve, but the latter has been attempted by Professor Hoffmann and Dr. Franz Grumbach in a paper presented to the Sixth European Conference of the International Association for Research in Income and Wealth, held at Portoroź, Yugoslavia, in August 1959

constructed from existing wholesale prices a new index which is made up of the individual indexes of wholesale prices for small groups of commodities, each weighted by its proportion to total household expenditure. Beginning in 1913, the official cost of living index is used. In this way the true growth of the economy has been indicated by deflating the money value of national income throughout the period under investigation.

4. ESTIMATES OF GERMAN NATIONAL INCOME, 1851–1957

The only areas in which an attempt has been made to determine national product generated have been agriculture and industry; the other sectors of the economy have been disregarded.[5] Previously, German national income had been estimated on the basis of distribution for the period from 1891 to 1913, although the Reich Statistical Office based its official study only on tax statistics of two member states of the German Reich.[6] The same office computed annual figures for distribution of national income during the interwar period and increasingly comprehensive figures for generation of national product in agriculture and industry. Since 1948 the West German Federal Statistical Office has made annual estimates of national income and its most important components in the German Federal Republic. In 1957 these figures were recomputed and the results used in the Hoffmann-Müller study under discussion.[7]

The methodological bases of the investigation are the concepts and definitions of income streams used in the standardized system of social accounts of the O.E.E.C.[8] In this system the principal streams of distribution of national income are: (1) employee's compensation; (2) entrepreneurial and property income accruing to households and private nonprofit organizations; (3) corporate savings; (4) direct corporate taxes; (5) entrepreneurial and property income of government, less (6) interest on the public debt. The main currents of national income are estimated for the changing area of Germany during the entire period 1851–1957. In addition, the principal streams and their com-

[5] Cf. Paul Jostock, *The Long-Term Growth of National Income in Germany*, Series V, *Income and Wealth* (London: 1955), pp. 79 ff.

[6] *Das deutsche Volkseinkommen vor und nach dem Kriege*, Einzelschrift zur Statistik des Deutschen Reiches, No. 24 (Berlin, 1932).

[7] H. Bartels, K. H. Raabe, O. Schörry, et al., "Die Neuberechnung des Sozialprodukts für die Bundesrepublik Deutschland," *Wirtschaft und Statistik*, N. F., Vol. 9 (1957), pp. 123 ff.

[8] Organization for European Economic Cooperation, *A Standardized System of National Accounts* (Paris: 1952), pp. 19, 75 ff.

ponents are given in detail for the following individual German states: Prussia, Saxony, Hesse, Baden, Württemberg, Bavaria, and for the city-states of Hamburg and Bremen for the period 1851–1913.

The statistical material available for the different streams of national income is extremely heterogeneous. This is particularly true of income tax data, since until 1920 income tax legislation was left in the hands of the individual states. For this period, the computation of income streams for the entire German Reich is based on estimates for the individual states. For the period until 1870, indeed, only results of Prussian income statistics are available; after that date, figures for other states are found increasingly. National income of the German Reich from 1851 to 1913 is estimated on the basis of per capita figures which are available for particular regions. It is assumed, obviously, that the results for the individual states are representative of the entire area; differences in income level resulting from the inclusion of data on federal states other than Prussia are adjusted after 1871. After 1920, national income is estimated on the basis of tax assessment data of the Reich.

Tax collection statistics of the federal states until 1913, and, since 1925, those of the Reich, are used to estimate entrepreneurial and wage and salary income of private households.[9] Before 1891, when Prussia reformed its Income Tax Law, some components of income had not been included and made subject to tax; estimates for the period before 1891 had to be adjusted by an "Unterbewertungszuschlag," i.e. an increase to take account of original undervaluation. Undistributed corporate profits are not included in tax statistics and had to be estimated in special studies. In contrast to the system of standardized accounts of the O.E.E.C., direct taxes paid by corporations are not separately computed to avoid double counting, since these taxes are paid out of undistributed profits and thus are included in the total figure. Data on government entrepreneurial and profit income, as well as interest on the public debt, are obtained from fiscal statistics and annual reports of social insurance agencies.

During the past century Germany has undergone many territorial changes, so that it is not possible to use the same geographical area over the entire period under investigation. This, of course, makes it difficult to use different periods for purposes of comparison.

[9] The period between 1914 and 1924 has been omitted; the wartime and post-war inflation do not yield reasonable figures which fit into the long-term trend.

5. RESULTS OBTAINED IN THE HOFFMANN-MÜLLER STUDY

All these problems, which are here mentioned only briefly, are the subject of rigorous scientific criticism in the Hoffmann-Müller study. This alone would render the book important, but it is, in addition, an excellent analysis of long-run national income computation. We shall not attempt to present all of the many rewarding results of the book but shall merely set forth a few of its important findings.

The development of nominal and real German national income between 1851 and 1957 is given in five-year averages, which affords a sufficient degree of precision for long-run analysis. The conclusions for single years may be dubious, but the correctness of the emerging trend is not affected. For an exact analysis of business fluctuations and cycles one must go back to annual values. These, however, are by their very nature not precise, since the most important components of income are considered in three-year-moving-averages in order to dampen oscillations and postpone cyclical turning points.

The expansion of population throughout the entire period is covered by computing per capita income. From 1851 to 1957 nominal national per capita income increased by more than ten times the initial value. In order to estimate changes in real per capita income the authors use their newly constructed cost of living index to deflate nominal income. Even with this adjustment real per capita national income rose to four times the original figure.

The period from 1851 to 1913 is of special interest, since it indicates the rapidity of economic growth with the increase in industrialization. In Germany the process of industrialization did not begin before the middle of the nineteenth century, while in England and France it started considerably earlier. The analysis in this book, therefore, provides a picture of industrial development almost from its very beginning. This is documented by the very low per capita income in the 1850's. From that time on, the per capita income rose quickly and steadily. Between 1851 and 1913 the population increased by some 80 percent, while the addition to total national income amounted to about 390 percent. That is, while population doubled and national income increased fivefold, the increase of per capita nominal income was about 170 percent, or almost triple. If one takes further account of the development of prices during this period, it is found that per capita real income is continuously above nominal income. Only the "Gründerjahre" (years of wild speculation) of 1871 to 1875 present an exception to this general conclusion. The period from 1851 to 1913 represented a steady increase of real per capita income, partly because

of a rise in nominal income and partly because of falling prices (up to 1900). After prices began to rise, the purchasing power of nominal income decreased, but real income continued to rise.

When we evaluate these findings, it must be remembered that changes in the relationship between the total population and the active working population have not been taken into account. Income per capita for the total population is, therefore, never identical with income per capita for the working population, nor is there a constant ratio between them, in view of the substantial increase in the ratio of the labor force to the total population since the first World War. The only conclusion that can be drawn is that real per capita income of the total population has roughly doubled in each sub-period of fifty years.

TABLE 1

NOMINAL AND REAL NATIONAL INCOMES IN GERMANY

Five-year Periods	Population (in thousands)	Nominal National Income		Price Index (1913 = 100)	Real National Income	
		Total (millions of marks, RM/DM)	Per Capita (marks, RM/DM)		Total (millions of marks, RM/DM)	Per Capita (marks, RM/DM)
1851–55	35,944	9,569	266	90.2	10,609	295
1856–60	36,884	10,755	292	96.3	11,168	303
1861–65	38,773	11,894	307	94.3	12,613	326
1866–70*	40,141	13,051	325	96.7	13,372	333
1871–75	41,641	15,171	364	103.4	14,672	352
1876–80	44,105	16,280	369	92.1	17,676	401
1881–85	46,039	17,557	381	82.0	21,411	465
1886–90	48,176	20,104	417	81.3	24,728	513
1891–95	50,825	22,638	445	80.2	28,227	555
1896–1900	54,405	27,028	497	79.4	34,040	626
1901–05	58,612	31,548	538	89.2	35,368	603
1906–10	62,863	39,919	635	92.5	43,156	686
1911–13	66,161	47,374	716	98.3	48,193	728
.....	.....	.....	.....	....	.....	...
1925–29	63,221	64,885	1,026	147.5	43,990	696
1930–34	64,928	49,849	768	128.8	38,703	596
1935–39	67,979	70,215	1,033	124.9	56,217	827
.....	.....	.....	.....	.....	.....	...
1950–54	47,825	99,297	2,076	209.0	47,511	993
1955–57	49,825	148,277	2,976	221.1	67,036	1,346

Source: W. G. Hoffmann, J. H. Müller, et al., *Das deutsche Volkseinkommen, 1851–1957* (Tübingen, 1959), p. 14. The territory covered changes.

* Average of the years 1866 and 1869.

Cross section analysis reveals that private households' share of total national income always amounts to more than nine-tenths, while undistributed corporate profits and government entrepreneurial and property income were quite insignificant at the beginning of the period. The percentage of national income going to the latter two recipients rose considerably, since their incomes grew more rapidly than those of private households. This is also true of interest on public debt, which in the process of industrialization increased extraordinarily with the greater role played by the state. This growth of interest on public debt was twice interrupted as a consequence of currency reform at the end of each World War.

Nominal income values are given by state from 1851 to 1913, but corresponding real figures cannot be computed because of lack of regional price indexes. In addition, the states investigated do not form definite and distinct economic regions, nor can they be characterized as predominantly industrial or agrarian. Very often the differences within a single state are of more importance than structural variations between states. A more precise study of regional differences in welfare and wealth (real income) will become possible only when smaller regional units are taken as the base for the analysis of national income, i.e., if the economic structure of the region is examined to detect differences in the composition of regional income. The very valuable study of Professors Hoffmann and Müller and their collaborators is a first step in this direction.

School of Advanced International Studies,
The Johns Hopkins University, Bologna, Italy

ECONOMETRICA

Journal of the Econometric Society

Vol. 28, No. 3 **July, 1960**

CONTENTS

JAMES DUESENBERRY, OTTO ECKSTEIN, and GARY FROMM: Stability and Instability in the American Economy.

A. H. LAND and A. G. DOIG: An Automatic Method of Solving Discrete Programming Problems.

RALPH E. GOMORY and WILLIAM J. BAUMOL: Integer Programming and Pricing.

I. RICHARD SAVAGE and KARL W. DEUTCH: A Statistical Model for the Gross Analysis of Transaction Flows.

A. L. NAGAR: A Monte Carlo Study of Alternative Simultaneous Equation Estimators.

GREGORY C. CHOW: Tests of Equality between Sets of Coefficients in Two Linear Regressions.

LIONEL W. MCKENZIE: A Contribution to the Theory of Competitive Equilibrium.

W. V. CANDLER: A Short-Cut Method for the Complete Solution of Game Theory and Feed-Mix Problems.

CHARLES W. HOWE: An Alternative Proof of the Existence of General Equilibrium in a von Neumann Model.

KENNETH J. ARROW and LEONID HURWICZ: Some Remarks on the Equilibria of Economic Systems.

Published Quarterly Subscription rate available on request

The Econometric Society is an international society for the advancement of economic theory in its relation to statistics and mathematics.

Subscriptions to ECONOMETRICA and inquiries about the work of the Society and the procedure in applying for membership should be addressed to Richard Ruggles, Secretary, The Econometric Society, Box 1264, Yale Station, New Haven, Connecticut, U.S.A.

The Quarterly Journal of Economics

Founded 1866

Vol. LXXIV **August 1960** **No. 3**

The Economics of "Bills Preferably" ..... Ralph A. Young and Charles A. Yager

Employment Growth and Price Levels: A Review Article ..Calvin B. Hoover

The Responses of the Bakers, Longshoremen and Teamsters to Public Exposure .....Philip Taft

A Note on Social Overhead and a Dynamic Analysis of Industry ..... Arthur H. Cole

Growth and Fluctuations: A Causal Interpretation .....Dale W. Jorgenson

Is Oligopoly Illegal? .....Jacob Weissman

Returns to Scale in Business Administration .....Martin J. Beckmann

Security and a Financial Theory of Investment .....Myron J. Gordon

Notes and Discussions

Economies of Scale: Some Statistical Evidence

Comment .....S. B. Alpert and S. C. Schuman

Reply .....Frederick T. Moore

Recent Publications

Edited by Arthur Smithies for the Department of Economics, Harvard University

$1.75 per copy; $6.00 a year; 3 years $17.00; Special Student Rate, $3.00 a year

Order from THE HARVARD UNIVERSITY PRESS

Cambridge 38, Massachusetts

THE JOURNAL OF POLITICAL ECONOMY

Edited by MARY JEAN BOWMAN, in Co-operation with the Other Members of the Economics Department of the University of Chicago

The August 1960 issue contains:

The Cost of Protection and the Scientific Tariff ..... *Harry G. Johnson*
Product Differentiation and Advertising: Some Lessons from Soviet Experience ..... *Marshall I. Goldman*
Differences Between the Personal Demand for Money and the Business Demand for Money ..... *John J. McCall*
Price-Level and Money-Denomination Movements ..... *John J. Klein*
Featherbedding—A Theoretical Analysis ..... *Paul A. Weinstein*
The Scene of the Crime of 1873 Revisited: A Note ..... *Paul M. O'Leary*
Kreinin, the Gatt Hypothesis, and Trade Diversion ..... *Steric T. Beza*
Returns to Speculators: Telser vs. Keynes ..... *Paul H. Cootner*
Reply ..... *Lester G. Telser*
Rejoinder ..... *Paul H. Cootner*
Book Reviews and Books Received

Published bimonthly in February, April, June, August, October, and December. Subscription rates for U.S.A. and possessions: 1 year, $6.00; 2 years, $11.50; 3 years, $16.50. Canada and Pan American Postal Union: Fifty cents more per year. All other countries: $1 more per year. Single copies; $1.75. Special Rate: $3.00 per year to full-time graduate students in economics or business.

The University of Chicago Press, 5750 Ellis Ave., Chicago 37, Ill.

AMERICAN ECONOMIC REVIEW

Volume L **June 1960** **Number 3**

ARTICLES

An Index of Soviet Industrial Output..N. M. KAPLAN and R. H. MOORSTEEN
Returns Equalization and Regional Growth ..... G. H. BORTS
Monetary Reserves and Credit Control ..... J. M. HENDERSON
The "Outer-Seven" and European Integration ..... M. E. KREININ
Western Development of the Rhodesias ..... STEPHEN ENKE
A Model of Price Flexibility ..... J. V. YANCE
The Allocation of Joint Costs ..... A. A. WALTERS

COMMUNICATIONS

Foreign Language Books in Economics ..... R. S. HOWEY
A Regional Countermeasure to National Wage Standardization ..... J. M. BUCHANAN and J. E. MOES
Population and Economic Growth:
Comment ..... H. H. VILLARD
Reply ..... E. E. HAGEN
The Shape of the Income Distribution:
Comment ..... F. G. ADAMS
Reply ..... STANLEY LEBERGOTT

THE AMERICAN ECONOMIC REVIEW, a quarterly, is the official publication of the American Economic Association and is sent to all members. The annual dues are six dollars. Address editorial communications to Dr. Bernard F. Haley, Editor, AMERICAN ECONOMIC REVIEW, Stanford University, Stanford, California. For information concerning other publications and activities of the Association, communicate with the Secretary-Treasurer, Dr. James Washington Bell, American Economic Association, Northwestern University, Evanston, Illinois. Send for information booklet.

OXFORD ECONOMIC PAPERS

Editorial Board: N. H. Leyland (General Editor), J. R. Hicks, C. M. Kennedy, Sir Donald MacDougall, G. B. Richardson, C. N. Ward-Perkins, P. P. Streeten (Secretary).

Volume 12 **February 1960** **Number 1**

CONTENTS

Economic Theory and Agrarian Economics.
.....By N. GEORGESCU-ROEGEN
Aggregated Rebate Schemes and Independent Competition.
..... By B. YAMEY
Notes on the Working of the Gold Standard before 1914.
..... By A. G. FORD
Earnings-Drift in the United Kingdom, 1948-57.....By J. MARQUAND
Cash-Flow in a Controlled Economy: the Case of Hungary.
..... By JANOS BARACS
A Note on the Measurement of Real Income.
..... By STANISLAW WELLISZ

Oxford Economic Papers is published three times a year in February, June and October. The pre-paid subscription price is 35s. 0d. post free, and the price of a single copy 16s. 0d. net. Orders may be placed with any bookseller or sent direct to the Publisher:

OXFORD UNIVERSITY PRESS

AMEN HOUSE, WARWICK SQUARE, LONDON, E.C.4.

SANKHYĀ

The Indian Journal of Statistics

Vol. 22 **Parts 1 & 2** **1960**

CONTENTS

Classification of Natural and Plantation Teak (Tectona Ghandis) Grown at Different Localities of India and Burma with Respect to Its Physical and Mechanical Properties .....K. R. Nair and H. K. Mukherji
A Brief History of the Organisation of Official Statistics in India during the British Period .....S. Subramaniun
Calculation of Sampling Errors for Index Numbers .....K. S. Banerjee
Type-Study on Peak Period in Harvesting Akan Paddy, West Bengal, December 1954–January 1955
.....R. D. Som, G. C. Bhattacharyya and N. K. Nzabeediri
Next Steps in Planning .....P. C. Mahalanobis
Industrialization of Underdeveloped Countries–A Means to Peace
..... P. C. Mahalanobis
Food Resources and Population of India–A Historical Study....P. C. Bannil
Test Scoring by the 602A Calculating Punch .....Ajit Haldar
Report: National Sample Survey Number Thirteen:
Report on Household Transport Operations
Corrigenda

For information please write to A. Mahalanobis, Editorial Secretary, SANKHYĀ, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta-35, India.

THE INTERNATIONAL ECONOMIC REVIEW is interested primarily in publishing articles on quantitative economics and welcomes contributions of empirical works, as well as those in mathematical economics and statistical theory related to quantitative aspects of economics. Quantitative economics will be broadly interpreted as covering both obviously quantitative studies and ideas that can eventually lead to quantitative studies. In spite of the emphasis on quantitative economics, the editors welcome also contributions to nonquantitative economics.

The Review is published fully in English, three times a year. The annual subscription price is $4.00 per year outside Japan and ¥1,200 per year in Japan; the price of a single copy is $1.50 outside Japan, ¥500 in Japan. Subscriptions outside of Japan should be forwarded to THE INTERNATIONAL ECONOMIC REVIEW, Wharton School of Finance and Commerce, University of Pennsylvania, Philadelphia 4, Pa., U.S.A. Subscriptions in Japan should be forwarded to THE INTERNATIONAL ECONOMIC REVIEW, Kansai Economic Federation, Osaka Building, Sozo-cho, Kita-ku, Osaka, Japan.

Manuscripts for publication should be sent to the chief editor or co-editor of the Review.

Editorial communication, books and periodicals for review, and communications respecting advertisements are to be sent to the business office of the Review.

SPRINGER-VERLAG IN WIEN

Nationalökonomie

Grundlagen für eine exakte Theorie

Von

Dipl.-Ing. **F. Korenjak**

Wien

Mit 84 Textabbildungen. XXIV, 374 Seiten. Gr.-8°. 1960

Ganzleinen $11.45

Die theoretische Nationalökonomie ist noch weit vom Status einer exakten Wissenschaft entfernt. Die vorliegende Arbeit hat sich die Aufgabe gestellt, ein Fundament für eine exakte Nationalökonomie zu errichten.

Die in früheren Zeiten verschiedentlich aufgestellte Behauptung, daß die Nationalökonomie nie eine exakte Wissenschaft werden könne, wird wohl kaum mehr ernst genommen und es fehlt nicht an Versuchen, eine derartige Theorie von verschiedenen Seiten in Angriff zu nehmen: Die „Ökonometrie" geht in analoger Richtung wie die exakten Naturwissenschaften vor, indem sie versucht, von den Erscheinungen induktiv auf die Gesetze des Wirtschaftens zu schließen (wie der Autor einleitend ausführt, kann dieser Methode nicht der gleiche Erfolg beschieden sein wie in den Naturwissenschaften, da in der Wirtschaft keine Versuche mit Variation von nur wenigen Größen angestellt werden können); rein mathematische Theorien stellen Axiomensysteme auf — eine Arbeit, die erst wertvoll wird, sobald man weiß, welche Axiomensysteme der theoretischen Nationalökonomie adäquat sind —; die Wiener Schule schließlich zeigte den Weg bzw. den Ausgangspunkt, wie der Autor im ersten Abschnitt darlegt.

Die unmittelbar evidenten Grundtatsachen des menschlichen Verhaltens werden als Grundlage der ganzen Theorie analysiert, und zwar wesentlich tiefer schürfend als von den Begründern und Nachfahren der Wiener und der Lausanner Schule. Davon ausgehend wird der Ablauf der Wirtschaft unter Verwendung einer möglichst einfachen Mathematik untersucht, wobei die unvermeidlichen idealisierenden Voraussetzungen in vieler Hinsicht weniger einschränkend sind als etwa bei der Lausanner Schule und ihren Nachfahren. Für eine Weiterentwicklung der Theorie unter schrittweiser Aufhebung dieser idealisierenden Voraussetzungen bildet eine Arbeit wie die vorliegende die unentbehrliche Grundlage, da nur so die begrifflichen Zusammenhänge erstmals geklärt und angemessene wirtschaftliche Begriffe exakt definiert werden können.

Zu beziehen durch Ihre Buchhandlung